



TELECOMUNICACIÓN

Campus Sur
POLITÉCNICA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y SISTEMAS DE TELECOMUNICACIÓN

PROYECTO FIN DE GRADO

TÍTULO: PREPROCESADO PARA LA MEJORA DE LA SEGMENTACIÓN
DEL ÁREA GLOTEL

AUTOR: Miguel Ros San Juan

TUTOR (o Director en su caso): Juana M^a Gutiérrez Arriola

TITULACIÓN: Electrónica de Comunicaciones

DEPARTAMENTO: Departamento de Teoría de la Señal y Comunicación

VºBº

Miembros del Tribunal Calificador:

PRESIDENTE: Fernando Pescador del Oso

VOCAL: Juana M^a Gutiérrez Arriola

SECRETARIO: Víctor José Osma Ruíz

Fecha de lectura: de de 20

Calificación:

El Secretario,

Agradecimientos

No resulta fácil explicar con palabras el agradecimiento que siento por las personas que me habéis brindado vuestro apoyo y vuestro conocimiento de forma tan desinteresada.

En este proyecto he necesitado contar con expertos en programación, en fisiología y anatomía médica, en tratamiento de imágenes, en redacción y estilo y, por supuesto, con todos aquellos que han ayudado a que consiga finalizar mis estudios en una carrera tan apasionante como esta.

Por eso me gustaría incluir en este agradecimiento a: mi tutora Juana, a Víctor, que se enfrentó por primera a este desarrollo, a mi madre, que ha sido capaz de aportarme el conocimiento necesario sobre la fisiología de las cuerdas vocales y por supuesto a todos aquellos autores que me han brindado la oportunidad de, a través de sus obras, sumergirme en el conocimiento del que trata este proyecto.

Resumen

A lo largo de este documento se presenta un estudio de distintas técnicas de preprocesado para mejorar la segmentación del espacio glotal en imágenes o vídeos de *análisis estrobolaríngoscópicos*.

La base de este trabajo la constituye la tesis doctoral publicada por el Dr. Ingeniero Víctor José Osma [1], en la que se realiza la segmentación. Este software de segmentación parte de la aplicación de la *transformada de Watershed* a una imagen - o fotograma de un vídeo - a la que previamente se le ha aplicado un gradiente en escala de grises. Posteriormente se realizan operaciones de fusión - *Merging* - de los objetos de la imagen para, finalmente, catalogar el elemento *glotis* mediante un discriminante lineal de *Fisher*.

Partiendo del software implementado, este trabajo pretende desarrollar una serie de algoritmos para mejorar la detección de la glotis.

La primera fase del proyecto será migrar la aplicación inicial a un entorno más moderno que facilitará tanto el actual trabajo como las posteriores aplicaciones.

Finalmente se realizará una segunda fase en la que se analizarán diferentes métodos de preprocesado de las imágenes con el fin de mejorar los resultados obtenidos. Con esto se pretende mejorar las características de las imágenes, distinguiendo mejor así los diferentes objetos presentes en la imagen y facilitando la labor de la distinción de la glotis.

Se han probado tres técnicas de preprocesado: normalización localizada, transformada discreta del coseno y filtrado anisotrópico. Obteniéndose mejores resultados con las últimas técnicas.

Abstract

This paper presents a study of several pre-processing techniques in order to improve the segmentation of the glottal area in images or in *strobe laryngocopy's analysis videos*.

This paper is based upon Engineer Víctor José Osma PhD dissertation [1], where segmenting is carried out. This segmentation software draws from the implementation of the *Watershed transform* into an image - or frame of a video - to which a grey scale gradient was previously applied. Merge operations of objects in the image - *Merging* - are subsequently carried out so as to eventually classify the glottis element through *Fisher* linear discriminant analysis.

On the basis of the implemented software, this work intends to develop a series of algorithms in order to improve the glottis detection.

The first stage in this project will consist of migrating the initial application into a more modern environment which will facilitate both the current work and subsequent applications.

Finally, a second stage will be carried out, where the different pre-processing methods of images will be analysed with the purpose of improving the obtained results. On the basis, the color of images is intended to be improved, thus distinguishing better the different objects appearing in the image and facilitating the work of glottis detection.

Three techniques have been tested: local normalization, discrete cosine transform and anisotropic filtering. Best results have been obtained with the latest.

Lista de acrónimos

RGB: *Red Green Blue*

JND: *Just Noticable Difference*

LN: *Local Normalization*

TF: *Transformada de Fourier*

DCT: *Discrete cosine transform*

IDCT: *Inverse Discrete cosine transform*

BMP: *Bit Maps Protocole*

JPEG: *Joint Photographic Experts Group*

AVI: *Audio Video Interleave*

IMG: *Abreviatura de Imagen*

EAX: *Extended Accumulator Register*

EBX: *Extended Base Register*

ECX: *Extended Counter Register*

Índice general

Agradecimientos	III
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.3. Estructura del documento	2
2. Antecedentes	5
2.1. Introducción	5
2.2. Fisiología del aparato fonador	6
2.2.1. Cavidad infraglótica	6
2.2.2. Cavidad glótica	7
2.2.3. Cavidad supraglótica	9
2.2.4. Patologías	10
2.3. Técnicas de análisis	12
2.3.1. Laringoscopia	12
2.3.2. Estroboscopia	14
2.3.3. Cinematografía	16

2.4. Software original	16
2.4.1. Gradiente	17
2.4.2. Transformada Watershed	20
2.4.3. "Merging" JND	25
2.4.4. "Merging" de regiones vecinas	27
2.4.5. Predictor lineal	27
3. Descripción de la solución propuesta	29
3.1. Introducción	29
3.2. Migración de código	29
3.2.1. Bloque imagen	30
3.2.2. Bloque reserva	31
3.2.3. Bloque inicio	32
3.3. Técnicas de preprocesado	33
3.3.1. Normalización Localizada (LN)	33
3.3.2. Transformada Discreta del Coseno	35
3.3.3. Filtro anisotrópico	38
4. Resultados	41
4.1. Introducción	41
4.2. Resultados de la migración del código	42
4.3. Resultados de las mejoras implementadas	43
4.3.1. Análisis de resultados para: LN	44
4.3.2. Análisis de resultados para: DCT	46

4.3.3. Análisis de resultados para: Filtrado anisotrópico	48
4.3.4. Análisis de resultados conjuntos	50
4.3.5. Análisis de resultados modificando parámetros	52
5. Manual de usuario	55
5.1. Pantalla principal	56
5.1.1. Explorador de archivos	56
5.1.2. Selección de parámetros	58
5.1.3. Rendimiento	59
5.2. Pantalla resultados	60
6. Consideraciones finales	63
6.1. Conclusiones	63
6.2. Líneas futuras	64
7. Anexos	65
Bibliografía	69

Índice de figuras

2.1. Aparato fonador [20]	6
2.2. Aparato fibroso [23]	7
2.3. Musculatura del aparato tensor [21]	8
2.4. Aparato motor [22]	8
2.5. Cavidad supraglótica [21]	9
2.6. Nódulos laríngenos [24]	10
2.7. Pólipos laríngenos [24]	11
2.8. Tumores laríngenos [24]	11
2.9. Laringoscopia indirecta	13
2.10. Laringoscopia directa	13
2.11. Ejemplo de ciclo captado mediante una estroboscopia	15
2.12. Ejemplo de ciclo captado mediante una estroboscopia	15
2.13. Esquema de la aplicación	16
2.14. Plantilla general 3x3 para el cálculo del gradiente	17
2.15. Plantillas para el cálculo del gradiente en imágenes: (1)Sobel, (2)Prewitt y (3)Roberts	18
2.16. Matriz inicial transformada de Watershed	20

2.17. Convenio de señalización utilizado en el algoritmo	21
2.18. Asignación de direcciones de la transformada de Watershed	22
2.19. Matriz transformada de Watershed una vez se han creado las cuencas. . . .	23
2.20. Matriz transformada de Watershed final	23
2.21. Ejemplo de transformada Watershed con gradiente (1) y con gradiente um- bralizado (2)	24
2.22. Umbral de visibilidad de los niveles de gris	25
2.23. Ejemplo de merging	26
2.24. Merging de regiones vecinas	27
3.1. Aplicación del algoritmo LN para diferentes tamaños de ventana	34
3.2. Aplicación del algoritmo de la DCT para distintos tamaños del lado del triángulo	37
3.3. Nomenclatura utilizada para el filtrado anisotrópico.	38
3.4. Ejemplos de imágenes aplicándoles el filtrado anisotrópico con distintos parámetros.	40
4.1. Resultados migración con errores	42
4.2. Resultados migración con mejoras	43
4.3. Resultados algoritmo LN: Glotis detectada	45
4.4. Resultados algoritmo LN: Glotis no detectada	45
4.5. Tabla de resultados DCT	46
4.6. Resultados DCT	47
4.7. Tabla de resultados filtrado anisotrópico	48
4.8. Resultados filtro anisotrópico	49

4.9. Tabla de resultados con ambos preprocesados	50
4.10. Resultados para ambos preprocesados	51
4.11. Resultados V0	52
4.12. Resultados V14	53
5.1. Pantalla principal de la aplicación.	56
5.2. Ejemplo de selección de archivo.	57
5.3. Pantalla resultados de la aplicación.	60
7.1. Resultados DCT	66
7.2. Resultados Filtro	67
7.3. Resultados Filtro + DCT	68

Capítulo 1

Introducción

1.1. Motivación

La voz es un vehículo de comunicación imprescindible en el día a día. Es por eso que se hace necesario un estudio sobre las posibles patologías que puedan surgir. En un principio este era un trabajo realizado exclusivamente por personal sanitario, en particular por *otorrinolaringólogos*. Hoy en día existen numerosas técnicas donde la tecnología entra a formar parte de estos estudios simplificando y facilitando el trabajo a estas personas.

Atendiendo al funcionamiento motor del sistema fonador - lugar en el que se genera la voz -, más concretamente a las cuerdas vocales, se pueden hacer diferentes estudios. Este trabajo se centrará en mejorar la segmentación del área glotal con el fin de detectar posibles anomalías en las cuerdas vocales, que afectan directamente a la voz.

Analizar el comportamiento de las cuerdas vocales es indispensable para el diagnóstico de estas patologías. Durante décadas se han ido desarrollando diferentes técnicas que facilitan este análisis. Pudiendo ser estos análisis de las señales acústicas producidas por el sistema fonador o mediante técnicas basadas en la visualización del mismo. Los métodos de visualización generalmente se centran en el área glotal, puesto que así se definen mejor las diferentes anomalías.

En 2010, el Dr. Ingeniero Victor José Osma Ruiz, presentó su tesis doctoral: *Contribución al procesamiento digital de imágenes para la caracterización de patologías laríngeas* [1]. Esta tesis está basada en el estudio y el análisis del área glotal. Teniendo como objetivo la detección de la glotis en los archivos de partida - bien imágenes o bien vídeos -.

Este trabajo fin de grado pretende continuar con dicha tesis aportándole diferentes mejoras con el fin de optimizar la detección de la glotis. Esta es, por tanto, la principal motivación.

1.2. Objetivos

Como se ha comentado anteriormente, este trabajo pretende actualizar la aplicación de partida [1], para conseguir los mejores resultados posibles. Para ello se hace necesario una planificación de objetivos que se comenta a continuación:

- Migración del software de la aplicación desarrollada en la tesis de partida. Este objetivo se plantea tanto para *actualizar* el entorno de programación, como para facilitar el posterior desarrollo de las mejoras a implementar. Para comprobar el resultado de la migración, se harán análisis con distintas imágenes.
- Desarrollo de diferentes preprocesados que aporten una mayor nitidez en el nivel de color de la imagen escogida por el usuario con el fin de mejorar los posteriores resultados.
- Realización de distintas pruebas para comprobar el correcto funcionamiento de la aplicación.

1.3. Estructura del documento

El trabajo ha quedado dividido en una serie de capítulos que se enumeran a continuación:

1. Introducción: Este capítulo pretende aportar una visión general sobre la motivación para la realización de este trabajo, así como la estructura del mismo y los objetivos planteados.
2. Antecedentes: en este capítulo se pondrá en antecedentes al lector. Está dividido en varias secciones. En primer lugar se comentará la fisiología del aparato fonador. Seguidamente se comentarán las posibles técnicas de análisis de dicho área, y por último se explicará el sistema de partida.
3. Descripción de la solución propuesta: en este capítulo se describe el trabajo realizado. Queda dividido en secciones, por un lado la migración del software y por otro lado las mejoras implementadas.

4. Resultados: en este capítulo se explicarán los resultados obtenidos para cada uno de los objetivos planteados.
5. Manual de usuario: este capítulo presenta un manual de usuario de la aplicación final.
6. Consideraciones finales: en este capítulo se comentan las conclusiones obtenidas, así como posibles vías de investigación para posibles trabajos futuros.

Capítulo 2

Antecedentes

2.1. Introducción

La voz es el vehículo de comunicación más importante que utiliza el ser humano. La voz juega un papel esencial en todos los ámbitos de nuestra vida para nuestra comunicación verbal y la comunicación de nuestros sentimientos, existiendo gran preocupación por tener y mantener un tono agradable de voz.

Es por eso que se hace necesario el estudio de este fenómeno con el fin de detectar posibles anomalías que alteren su correcto funcionamiento.

El lenguaje humano involucra de forma especial al aparato respiratorio, pero también comprende:

1. Centro del lenguaje en la corteza cerebral
2. Los centros respiratorios del tallo cerebral

Este estudio se va a centrar en la parte mecánica de aparato fonador.

2.2. Fisiología del aparato fonador

El aparato fonador tiene tres partes fundamentales (Figura: 2.1): las *cavidades infraglóticas* (órganos respiratorios), la *cavidad glótica* (órgano fonador) y las *cavidades supraglóticas* (órganos de la articulación) [2] [3].

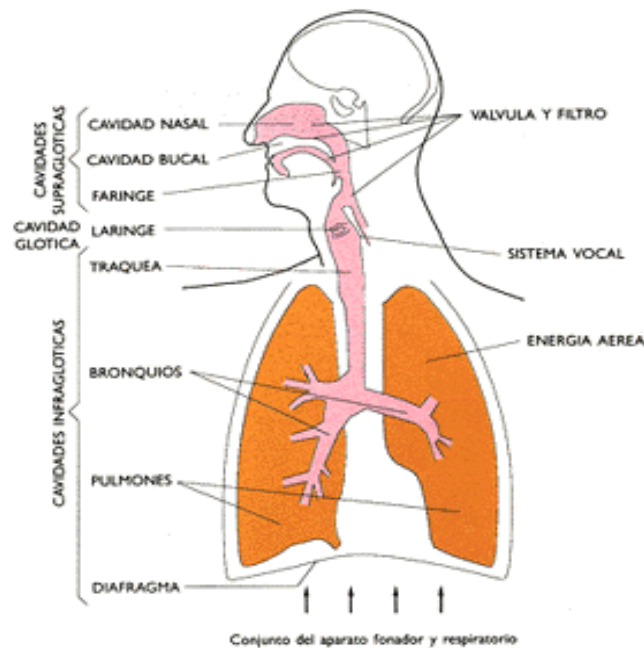


Figura 2.1: Aparato fonador [20]

2.2.1. Cavidad infraglótica

La cavidad infraglótica está ubicada justo por debajo de la glotis, y está compuesta de: *diafragma, pulmones, bronquios y tráquea*.

El diafragma es un tabique muscular que se sitúa entre la cavidad torácica y la abdominal. Cuando se contrae - al mismo tiempo que se elevan las costillas - aumenta el volumen de la cavidad torácica y el aire entra en los pulmones - *inspiración* -.

De la misma manera, cuando el diafragma se relaja provoca una expulsión del aire contenido en los pulmones con la fuerza y el ritmo necesario para la fonación - *expiración* -.

2.2.2. Cavidad glótica

La laringe está especialmente adaptada para actuar como un vibrador. Desde un punto de vista funcional comprende:

- El **aparato fibroso** (Figura: 2.2), soporte de las cuerdas vocales y su esqueleto: cartílagos (cricoide, tiroide, epiglótico, aritenoides, Santorini, Morgagni y sesamoideos), articulaciones y ligamentos

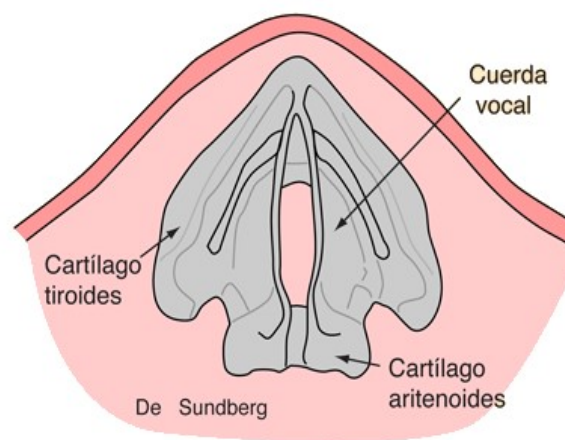


Figura 2.2: Aparato fibroso [23]

- El **aparato tensor de las cuerdas vocales** (Figura: 2.3). Los elementos vibradores son las cuerdas vocales, pliegues a lo largo de las caras laterales de la laringe que son estirados y movidos por los diferentes músculos específicos de la propia estructura laringea. Cada cuerda vocal es estirada entre el *cartílago tiroides* y el *aritenoides* por los diferentes músculos inversamente, la contracción de los músculos cricoaritenoides laterales tiran de los *cartílagos aritenoides* hacia adelante permitiendo la respiración normal. El músculo *Tiro-aritenoides* cuya capa interna, conocida como *músculo vocal*, constituye la mayor parte del grosor de la cuerda vocal, cuya consistencia y tensión incrementa a la vez que estrecha la glotis.

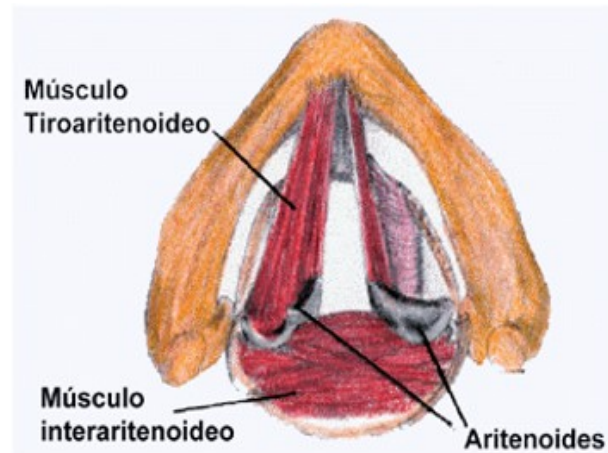


Figura 2.3: Musculatura del aparato tensor [21]

Las cuerdas vocales pueden adoptar diferentes formas, en función de las cuales cambia el tono así pues, unas cuerdas vocales gruesas producen sonidos graves mientras que unas planas o delgadas producen sonidos agudos.

- El **aparato motor de las cuerdas vocales** (Figura: 2.4). Lo constituyen los músculos que aproximan o separan las cuerdas vocales. Así, los *músculos crico-aritenoides* provocan una separación de las cuerdas vocales abriendo ampliamente la glotis durante la inspiración. Los *músculos cricoideos laterales* son sus antagonistas y provocan la aproximación de las cuerdas vocales y preparan la acción del *músculo vocal*. El *músculo aritenario transversal* tira de los aritenoides a la vez y produce una aproximación de las cuerdas vocales de forma que vibren con el aire espirado [4]

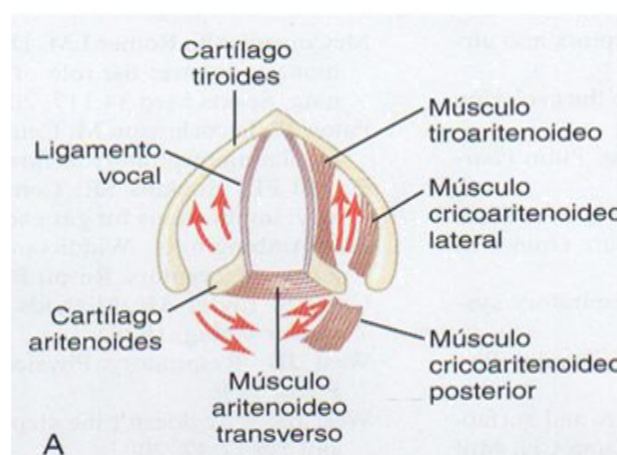


Figura 2.4: Aparato motor [22]

2.2.3. Cavity supraglótica

Son los órganos de articulación y resonancia. Los tres órganos principales de la articulación son los labios, la lengua y el paladar blando y los responsables de la resonancia incluyen la boca, la nariz, senos paranasales, faringe e incluso el torax. Es el último tramo que recorre la columna de aire, y donde se genera la gran variedad de sonidos gracias a la movilidad de la cavidad bucal.

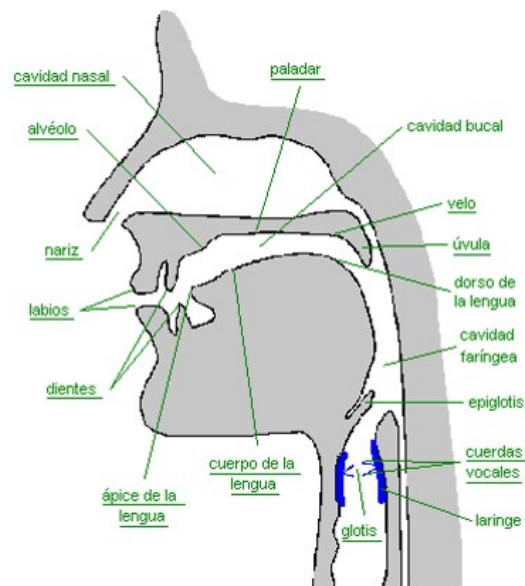


Figura 2.5: Cavidad supraglótica [21]

2.2.4. Patologías

El sistema fonador puede sufrir diferentes patologías, lo que implica una distorsión en la voz - voz patológica -. Para explicar este apartado, se van a diferenciar dos tipos de patologías. Por un lado, las patologías de tipo orgánico, que implican una lesión en el órgano principal de la voz, la laringe. Y por el otro patologías de tipo funcional, que son aquellas en las que la voz se ve afectada por un mal funcionamiento de la laringe [5] [6].

Patologías orgánicas:

1. Nódulos laríngeos: se producen generalmente en personas que realizan un sobreesfuerzo vocal durante largos periodos de tiempo. Esto produce una pequeña inflamación de la mucosa que recubre las cuerdas y por consiguiente genera vibraciones y rozamientos indebidos, lo que se traduce en alteraciones en el timbre de la voz.



Figura 2.6: Nódulos laríngeos [24]

2. Pólipos laríngeos: son protuberancias parecidas a callos o cicatrices que crecen alrededor de las cuerdas vocales debido a diferentes causas. Estos pólipos dificultan el paso del aire a través de la glotis, impidiendo así un correcto funcionamiento del sistema fonador.



Figura 2.7: Pólipos laríngeos [24]

3. Tumores laríngeos: Son masas tumorales de tamaño variable, afectan a la vía respiratoria. Generalmente se encuentra en el vestíbulo laríngeo y se originan a partir de elementos epiteliales o tejido conectivo. Comúnmente producen afonías o disfonías.



Figura 2.8: Tumores laríngeos [24]

4. Relacionadas con el sistema nervioso de la larínge, entre ellas:
- Espasmos e hiperquinesias.
 - Parálisis uni o bilateral de cuerdas vocales.
 - Disquinesias o alteración de la coordinación de los movimientos.

Patologías funcionales:

Son las que se producen por:

- Esfuerzos vocales o deficiente técnica fonatoria de las personas que trabajan con la voz.
- Debilidad neuromuscular laríngea hereditaria.
- Alteraciones debidas a problemas hormonales.
- Alteraciones psíquicas y afectivas como la afonía psíquica, atonía laríngea, etc.

2.3. Técnicas de análisis

Existen diversas técnicas de análisis del sistema fonador, unas permiten la visualización de las cuerdas vocales, mientras que otras permiten el registro de los mecanismos aerodinámicos, vibratorios o acústicos que tienen lugar en la laringe [1][9][10][11].

2.3.1. Laringoscopia

Consiste en la visualización de la laringe y las cuerdas vocales por medio de un laringoscopio. Este instrumento esta formado por un sistema óptico - bien un espejo o bien fibra óptica - y una fuente de luz. En función del tipo de laringoscopio utilizado distinguimos:

1.- Laringoscopia indirecta: Se realiza colocando un espejo, denominado espejo laríngeo, en la faringe o en la zona posterior del velo del paladar. Aplicando una fuente de luz, se ilumina la región laríngea obteniendo así una imagen invertida de la laringe. Por otro lado, el instrumento utilizado puede ser un pequeño telescopio. En este caso, éste se introduce por vía nasal hasta la garganta. Es la forma más común para examinar la laringe.

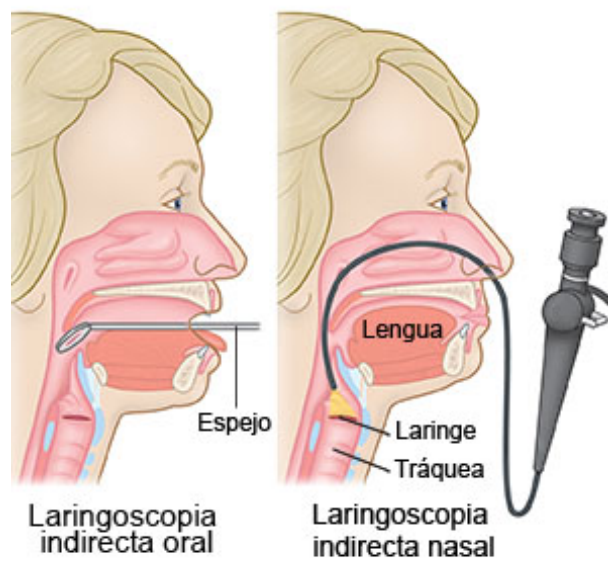


Figura 2.9: Laringoscopia indirecta

2.- Laringoscopia directa: Este tipo de análisis permite la visualización directa del área glotal. El laringoscopio se coloca en la parte posterior de la garganta, este método proporciona al médico una imagen más profunda y exacta de la garganta.

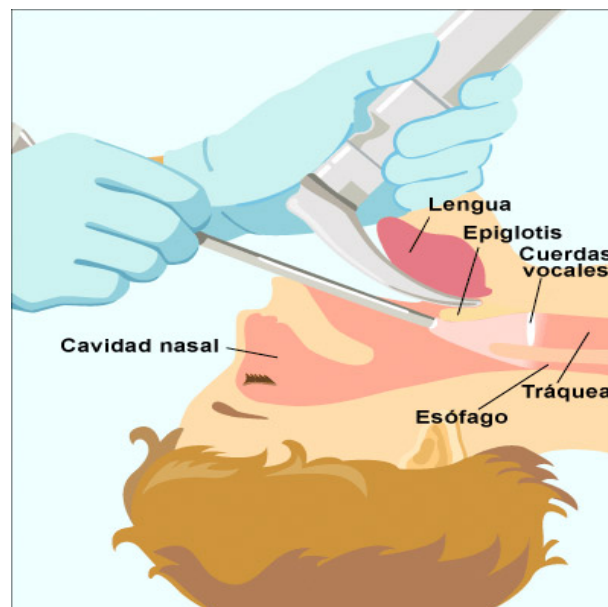


Figura 2.10: Laringoscopia directa

2.3.2. Estroboscopia

Este método ofrece la posibilidad de explorar el mecanismo de ondulación de la mucosa de las cuerdas vocales, y es el que se utiliza para la realización de este estudio.

De un modo resumido, se puede decir, que la estroboscopia es un estudio dinámico que permite observar la laringe en funcionamiento por medio de flashes intermitentes.

El estroboscopio es un instrumento inventado por Simon von Stampfer en 1829¹. Se utiliza para visualizar objetos en movimiento como si estuviera inmóvil. Este fenómeno se consigue encendiendo y apagando una luz intermitentemente.

El estroboscopio está dotado de una lámpara similar a la de un flash de una cámara fotográfica, que emite una serie de destellos consecutivos a una frecuencia regulable. Si tenemos un objeto girando a N revoluciones por minuto y regulamos la frecuencia a N destellos por minuto e iluminamos al objeto en cuestión, este aparecerá inmóvil a nuestra vista. De la misma forma, si la frecuencia no es la misma pero es aproximada, veremos al objeto moverse lentamente.

La vibración de los pliegues vocales durante la fonación es un fenómeno que bajo condiciones normales se produce a frecuencias muy altas - para los hombres de 120 a 140 Hz, para las mujeres de 200 a 220 Hz y en el caso de los niños puede llegar hasta los 280 Hz -. La luz estroboscópica proporciona una imagen de aparente lentitud de las vibraciones para que estas puedan ser analizadas.

La luz que emite el estroboscopio, es disparada a la misma frecuencia de vibración que las cuerdas vocales. Su objetivo es tomar imágenes individuales fijas en diferentes puntos de los ciclos vibratorios y así poder representarlas como un movimiento del pliegue vocal a una frecuencia mucho más baja.

En la figura 2.11 se representa un ejemplo de este fenómeno. La función $x(t)$ representa los periodos de vibración de las cuerdas vocales (ciclos de vibración). La función $p(t)$ indica los periodos de tiempo en los que se produce un destello de luz y por último, $y(t)$ representa la onda que se vería a simple vista.

¹Matemático e inventor austríaco nacido el 26 de octubre de 1792.

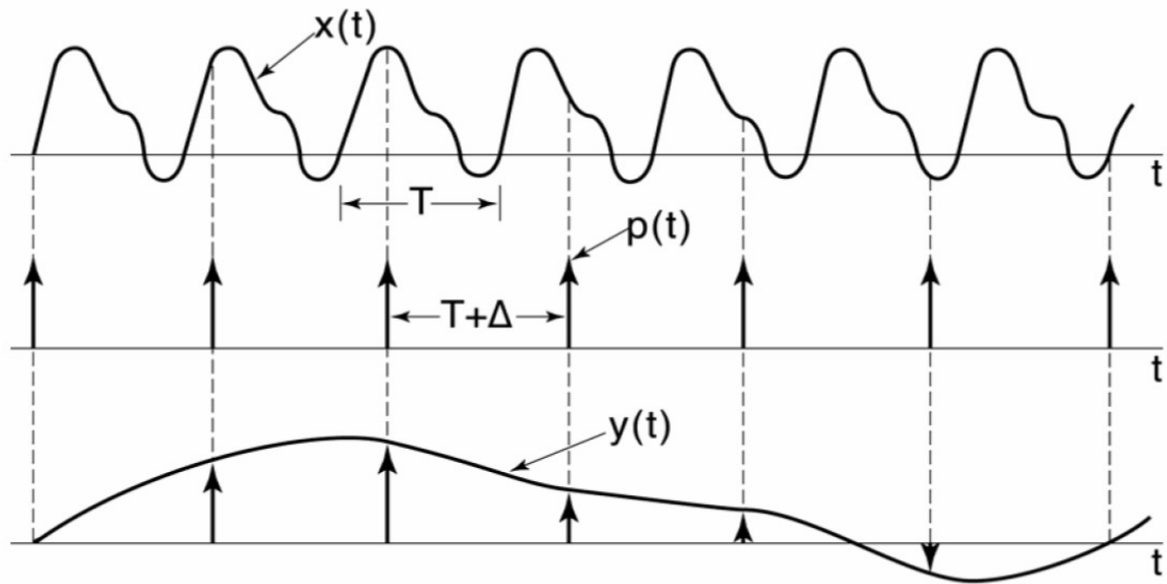


Figura 2.11: Ejemplo de ciclo captado mediante una estroboscopia

Se puede observar que mediante este método son necesarios 7 ciclos de vibración real para poder observar un ciclo completo a simple vista.

Finalmente, en la figura 2.12 se puede ver un ciclo completo de una estroboscopia:

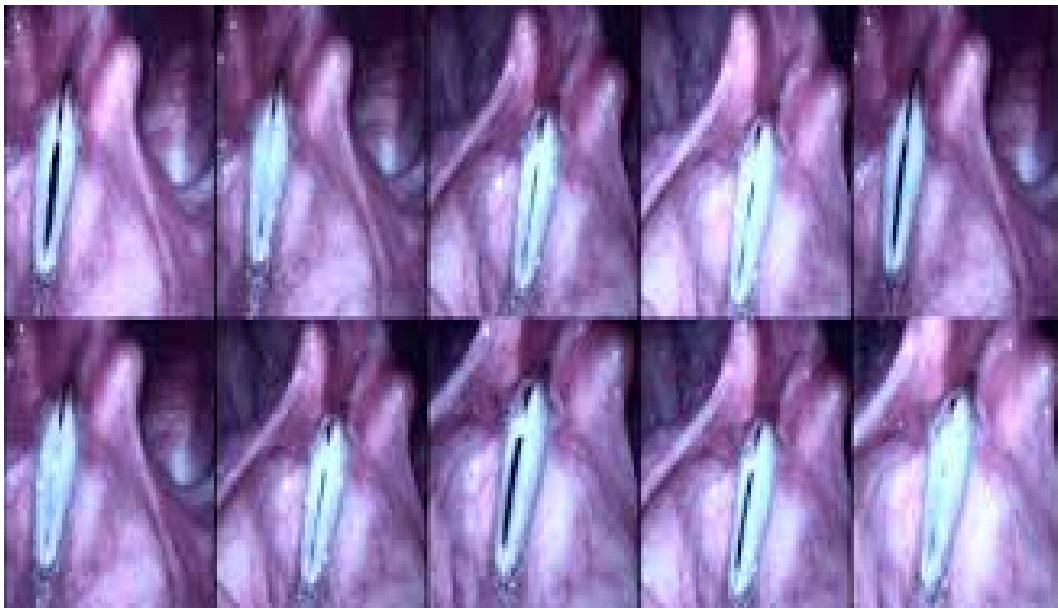


Figura 2.12: Ejemplo de ciclo captado mediante una estroboscopia

2.3.3. Cinematografía

Se trata de un sistema contrario a la *estroboscopia* que registra los fotogramas a una velocidad mucho mayor que en un vídeo convencional.

Las imágenes se monitorizan a una velocidad más lenta que permita ser visualizada por el ojo humano.

2.4. Software original

Para poder entender el funcionamiento de las mejoras de la aplicación, es necesario hacer un repaso de la aplicación inicial [1]. En la figura 2.13 se puede observar un esquema del funcionamiento de dicha aplicación:

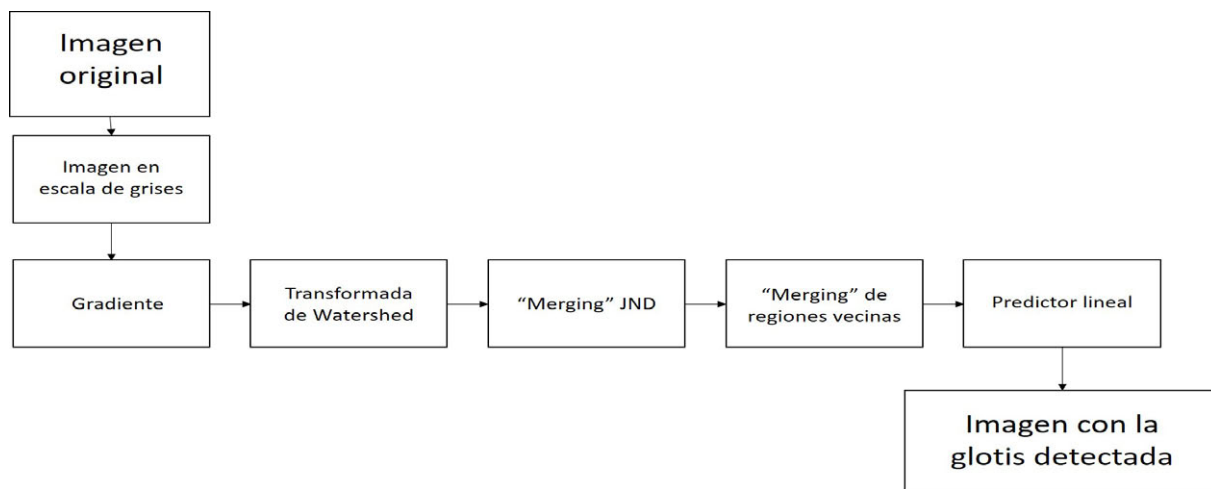


Figura 2.13: Esquema de la aplicación

Como se puede ver en la imagen, se parte de una imagen en color, y posteriormente se le aplican una serie de algoritmos para conseguir la detección de la glotis. El resultado final es una imagen en escala de grises con la glotis detectada.

El primer paso es transformar la imagen inicial en escala de grises. Una vez se ha transformado la imagen, se procede a realizar cada uno de los bloques indicados anteriormente, que se van a explicar a continuación:

2.4.1. Gradiente

Este es el primer paso del sistema inicial, es una forma de delimitar objetos detectando sus bordes. Se aplica sobre la imagen inicial en escala de grises, para obtener mejores resultados a la hora de detectar el área global.

En sí mismo, el gradiente es el resultado de derivadas a lo largo de los ejes de abscisas y de ordenadas respectivamente. El cálculo del gradiente se obtiene mediante las ecuaciones 2.1 y 2.2:

$$G_x = \frac{\partial f(x, y)}{\partial x} \quad G_y = \frac{\partial f(x, y)}{\partial y} \quad (2.1)$$

$$|G(x, y)| = \sqrt{G_x^2 + G_y^2} \quad (2.2)$$

En el caso de calcular el gradiente en imágenes, éste se obtiene mediante plantillas por medio de las cuales se calcula la diferencia de luminancia entre el píxel estudiado - colocado en el centro de la plantilla - y sus vecinos. En la figura 2.14 se puede ver un ejemplo de plantilla general de tamaño 3x3:

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6
Z_7	Z_8	Z_9

Figura 2.14: Plantilla general 3x3 para el cálculo del gradiente

Para el cálculo del módulo del gradiente según esta plantilla se hace uso de la ecuación 2.3:

$$|G| = \sqrt{[(x_3 + x_6 + x_9) - (x_1 + x_4 + x_7)]^2 + [(x_7 + x_8 + x_9) - (x_1 + x_2 + x_3)]^2} \quad (2.3)$$

Con cada uno de los factores calculado mediante la ecuación 2.4, siendo f_i el valor del pixel y z_i su posición en la plantilla:

$$x_i = z_i \cdot f_i \quad (2.4)$$

En el software inicial se implementan tres métodos: *Sobel*, *Prewitt* y *Roberts* basados a su vez en diferentes plantillas. Los dos primeros implementan dos plantillas, una para el cálculo del gradiente del eje horizontal y otra para el cálculo del gradiente en el eje vertical, mientras que el tercero utiliza dos plantillas mezclando cálculos de derivadas, sin estar éstas sujetas a un eje determinado. En la figura 2.15 se pueden distinguir estos 3 métodos. La primera columna se corresponde con las dos plantillas del método de Sobel, la segunda con las del método de Prewitt y la última con el método de Roberts.

<table> <tr><td>-1</td><td>-2</td><td>-1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>2</td><td>1</td></tr> </table>	-1	-2	-1	0	0	0	1	2	1	<table> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	-1	-1	-1	0	0	0	1	1	1	<table> <tr><td>0</td><td>0</td><td>-1</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	0	0	-1	0	1	0	0	0	0
-1	-2	-1																											
0	0	0																											
1	2	1																											
-1	-1	-1																											
0	0	0																											
1	1	1																											
0	0	-1																											
0	1	0																											
0	0	0																											
<table> <tr><td>-1</td><td>0</td><td>1</td></tr> <tr><td>-2</td><td>0</td><td>2</td></tr> <tr><td>-1</td><td>0</td><td>1</td></tr> </table>	-1	0	1	-2	0	2	-1	0	1	<table> <tr><td>-1</td><td>0</td><td>1</td></tr> <tr><td>-1</td><td>0</td><td>1</td></tr> <tr><td>-1</td><td>0</td><td>1</td></tr> </table>	-1	0	1	-1	0	1	-1	0	1	<table> <tr><td>-1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	-1	0	0	0	1	0	0	0	0
-1	0	1																											
-2	0	2																											
-1	0	1																											
-1	0	1																											
-1	0	1																											
-1	0	1																											
-1	0	0																											
0	1	0																											
0	0	0																											
(1)	(2)	(3)																											

Figura 2.15: Plantillas para el cálculo del gradiente en imágenes: (1)Sobel, (2)Prewitt y (3)Roberts

Además de los métodos de detección de bordes comentados, existen otros dos. El primero de ellos está basado en el método de la segunda derivada, se usa en imágenes en las que los bordes de los objetos que la componen no son tan abruptos y es más sensible al ruido presenta en las imágenes, en este caso no presentaría buenos resultados. Y el segundo es un método más avanzado, el *Algoritmo de Canny*. Este método presenta resultados mucho mejores que los anteriores pero disminuye considerablemente el rendimiento.

El método utilizado en la aplicación inicial es el de Sobel, puesto que el método de *Canny* afecta al rendimiento y es menos sensible al ruido existente en las imágenes que los métodos de *Roberts* o *Prewitt*.

Es muy común que las imágenes no tengan niveles de gris homogéneos, por lo que se hace necesario *umbralizar* el gradiente. Este método consiste en aplicar un umbral - 2 en el caso del software implementado - mediante el cual, todos los píxeles con gradiente de valor menor que dicho umbral, toman valor cero. Mientras que los demás píxeles mantienen su valor. Matemáticamente es muy fácil de calcular, y tomando V como el valor final del píxel y x como el valor de este, se podría expresar según la ecuación 2.5:

$$V = \begin{cases} 0 & \text{si } x > 2 \\ x & \text{si } x \leq 2 \end{cases} \quad (2.5)$$

Aplicando este método se consiguen eliminar *falsos bordes* producidos por el ruido o por una mala iluminación. Así, se consigue reducir el número de regiones en un 20 % aproximadamente.

2.4.2. Transformada Watershed

Esta es una herramienta que permite dividir la imagen inicial en diferentes objetos en función de alguna característica, como puede ser su nivel de gris o el valor del gradiente. Al ser una imagen en escala de grises, los valores que pueden tomar los píxeles estarán comprendidos entre 255 y 0. Siendo 0 el color negro y 255 blanco.

Como su nombre indica, este método divide la imagen en diferentes cuencas, es decir, zonas con nivel de gris similar. Asemejándose así al efecto que tienen las gotas de agua cuando llueve, si éstas caen en lo alto de una superficie, caerán hasta agruparse en distintas cuencas. Estas cuencas estarán formadas por los *mínimos regionales*, que son en definitiva, un pixel o conjunto de píxeles con el menor nivel de gris de entre sus vecinos.

Una imagen en escala de grises es una matriz de dos dimensiones, en la que cada “celda” es un píxel. El objetivo de este método es recorrer la matriz inicial y generar una matriz de iguales dimensiones en la que aparezcan “cuencas”. Cada uno de los píxeles de la matriz inicial únicamente podrá pertenecer a una de las cuencas.

Para entender mejor este método se analizará una matriz sencilla, en la que cada una de las celdas representa un pixel y su valor se corresponde con el nivel de gris de dicho pixel:

La matriz inicial, normalmente, será una imagen a la que se le ha aplicado el gradiente, ya que así se identifican mejor los bordes de cada uno de los objetos que componen la imagen.

195	194	193	194	195	193	253	195	188	254	0
194	126	194	195	194	254	194	250	131	0	254
192	130	125	195	189	122	128	128	0	254	252
194	191	196	193	197	129	0	0	192	253	252
254	194	194	194	124	1	0	128	196	188	254
253	194	197	121	4	0	121	196	199	194	253
196	124	197	128	0	121	200	194	189	254	254
126	126	193	0	195	199	192	0	194	252	254
193	3	0	193	194	253	195	196	192	254	253
195	0	192	254	254	192	192	194	198	247	254
0	254	254	254	251	254	254	254	249	254	252

Figura 2.16: Matriz inicial transformada de Watershed

El primer paso consiste en recorrer la matriz inicial y marcar los mínimos - píxeles con menor valor de gradiente -. Por otro lado, el algoritmo, irá asignando la dirección en la que se encuentra el píxel vecino con menor valor de gradiente - siempre y cuando no sea una cuenca - mediante flechas, como podemos ver en la figura 2.18 .

Para asignar dichas flechas, el algoritmo se posiciona en un píxel y captura el valor de gradiente de todos y cada uno de sus píxeles vecinos, detectando el menor de ellos. Según la dirección de la flecha, el algoritmo asigna un número al píxel analizado. En la figura 2.17 se ve una relación de las flechas con el número al que asigna:

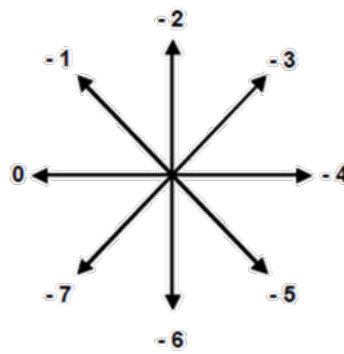


Figura 2.17: Convenio de señalización utilizado en el algoritmo

Los números seleccionados para etiquetar a cada píxel en función de la flecha, deben ser negativos, ya que los valores que contienen los píxeles están comprendidos entre 0 y 255. Así, por un lado va señalando los píxeles analizados, y por otro, que los valores asignados no se confunden con valores del gradiente .

En la siguiente imagen se puede ver el resultado una vez se ha analizado la matriz de ejemplo que se muestra en la figura 2.18:

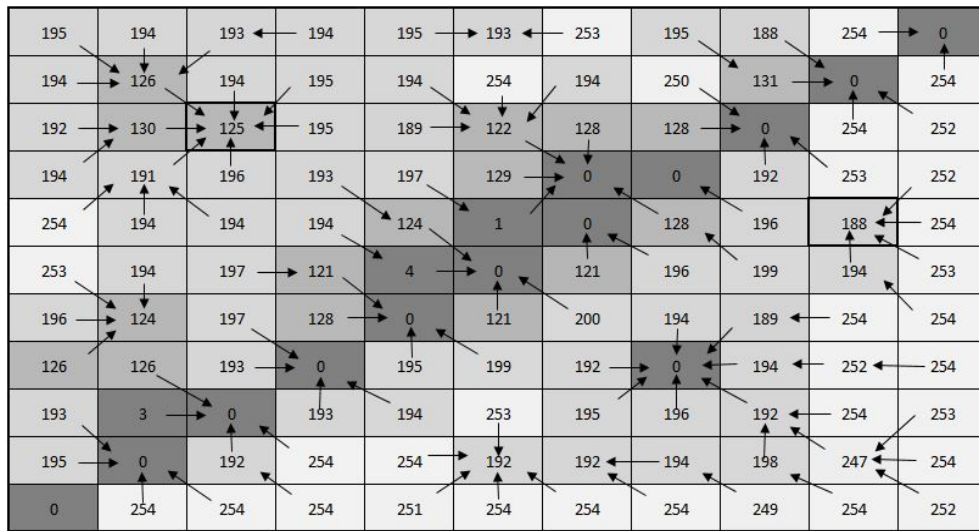


Figura 2.18: Asignación de direcciones de la transformada de Watershed

Cuando ya el algoritmo ha recorrido todos los píxeles y les ha asignado un valor según la dirección del píxel vecino con menor valor, recorre de nuevo la matriz. Los mínimos estarán marcados con un valor entero positivo, por lo que en este paso, el algoritmo se posiciona en ellos y recorre los píxeles vecinos que tengan la flecha apuntando hacia él, pero en sentido contrario, de tal forma que cada píxel queda etiquetado en una cuenca. Para calcular el camino que sigue la *gota de lluvia*, se hace uso de la *distancia geodésica*. Este método detecta la menor distancia que se debe recorrer para alcanzar el punto de destino. En este caso es el camino hasta los mínimos marcados.

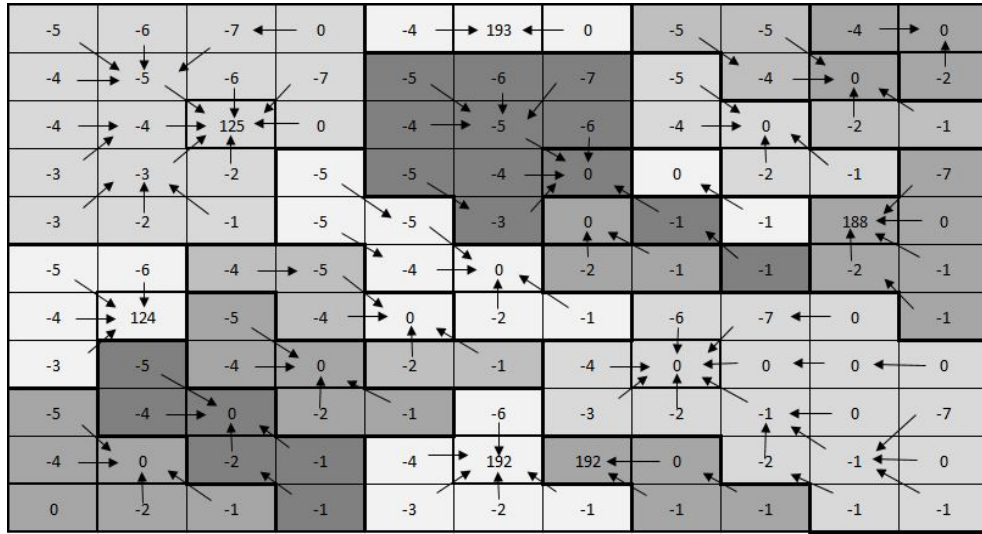


Figura 2.19: Matriz transformada de Watershed una vez se han creado las cuencas.

Por último, se marca a cada píxel con el valor del mínimo de la cuenca a la que pertenece, de tal forma que quedan definidas las cuencas. En la siguiente imagen se puede ver el resultado final de la matriz transformada de Watershed.

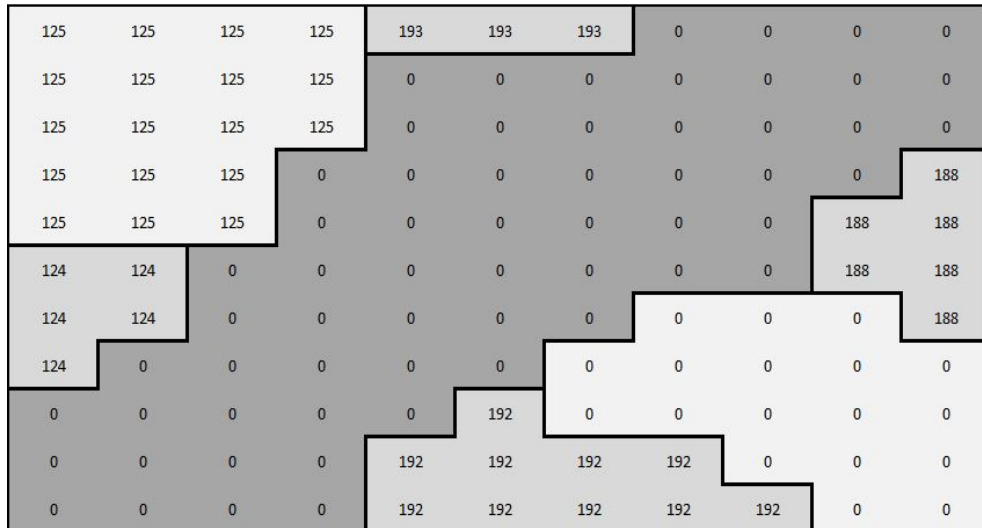


Figura 2.20: Matriz transformada de Watershed final

En la figura 2.21 se puede observar la aplicación de la transformada Watershed sobre dos imágenes. A la primera de ellas se le ha aplicado un gradiente mientras que a la segunda se le ha aplicado un gradiente umbralizado (2.4.1).

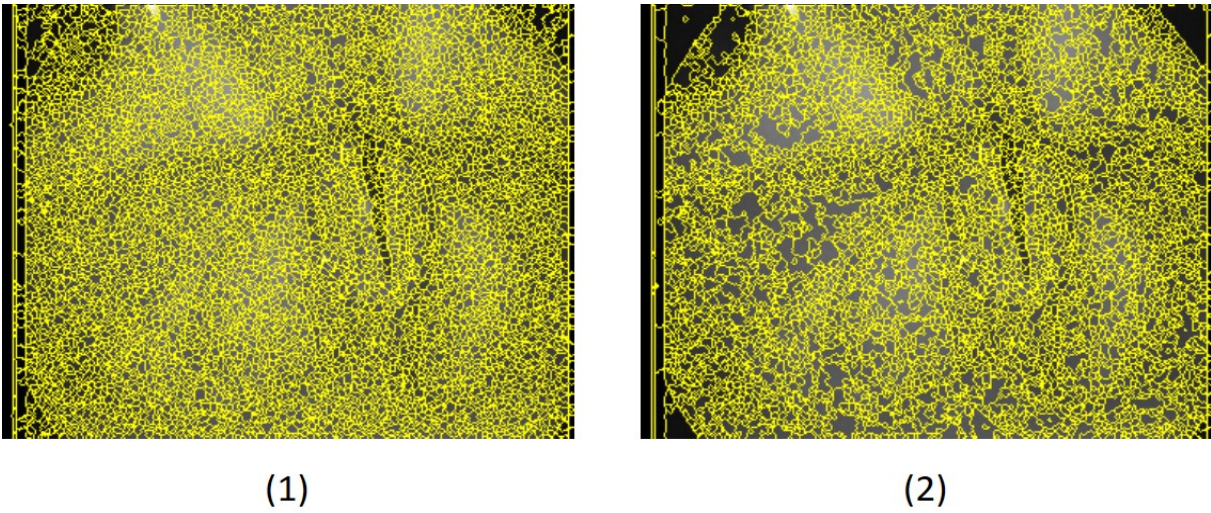


Figura 2.21: Ejemplo de transformada Watershed con gradiente (1) y con gradiente umbralizado (2)

2.4.3. "Merging" JND

Como se puede ver en la figura 2.21 aplicando únicamente la transformada de Watershed, quedan definidas un gran número de cuencas, por lo que se hace imposible la detección de la glotis. Por eso se hace uso de operaciones de *Merging*.

Esta herramienta trata de fusionar las cuencas con un nivel de gris similar utilizando la técnica de "Just Noticeable Difference" - de ahora en adelante JND - El JND es una medida que indica cuanto ha tenido que cambiar algo, para que este cambio sea apreciable. En nuestro caso, se aplicará a los distintos niveles de gris de cada una de las cuencas, pudiendo así fusionar aquellas cuencas vecinas con un nivel de gris diferente que no es apreciable por el ojo humano.

Gráficamente, este efecto se podría representar mediante la gráfica 2.22:

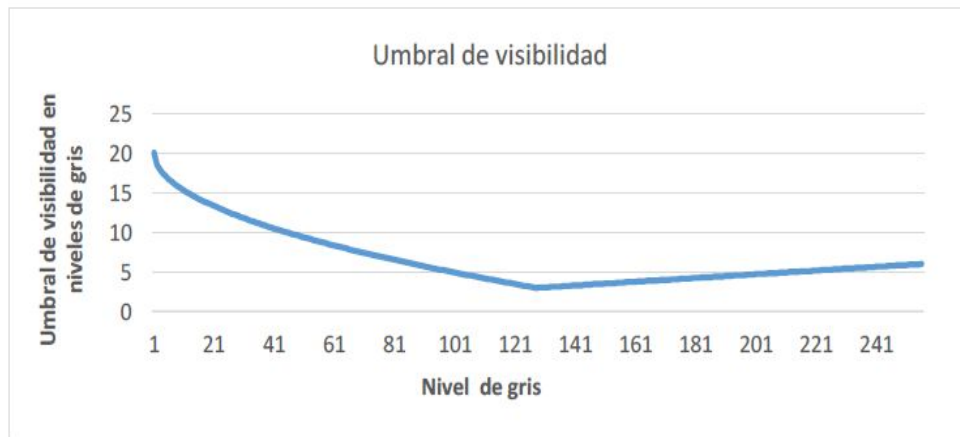


Figura 2.22: Umbral de visibilidad de los niveles de gris

La representación matemática del umbral de visibilidad T mostrado en la gráfica anterior quedaría definido según la ecuación 2.6:

$$T(n, m) = \begin{cases} 17 \cdot (1 - \sqrt{\frac{I(n, m)}{127}}) + 3 & \text{si } I(n, m) \leq 127 \\ \frac{3}{128} \cdot (I(n, m) - 127) + 3 & \text{resto} \end{cases} \quad (2.6)$$

La función de coste utilizada se calcula según la siguiente expresión:

$$F_c = [|mR_1 - mR_2| - MinJND(mR_1, mR_2) + 255] \cdot \frac{MinArea(R_1, R_2)}{LimitArea} \quad (2.7)$$

Donde mR_i equivale al valor medio de gris de cada cuenca R_i y $LimitArea$ establece un valor límite de área por debajo del cual se facilita la unión en caso contrario se dificulta.

Esta función de coste devuelve valores superiores a 255 para niveles de gris que el ojo humano es capaz de distinguir, por lo tanto por debajo de dicho umbral, se permite la unión de las cuencas.

Por ejemplo, para dos cuencas con niveles de gris 45 y 50 respectivamente. La mínima JND se conseguirá, a partir de la expresión 2.6, para la cuenca con valor de gris 50 y tendrá un valor de 9.33. Esto hace que la cuenca con valor 45 sea indistinguible para el ojo humano, puesto que $50 - 9,33 < 45$. El valor que devolvería la función de coste sería de 250.67, que está por debajo de 255, por lo tanto se permitiría la unión de las dos cuencas.

En la figura 2.23 se puede ver un ejemplo real. La primera imagen es la imagen inicial antes de procesarla, y en la segunda se pueden observar las cuencas que quedan sin fusionar una vez hecho este proceso.

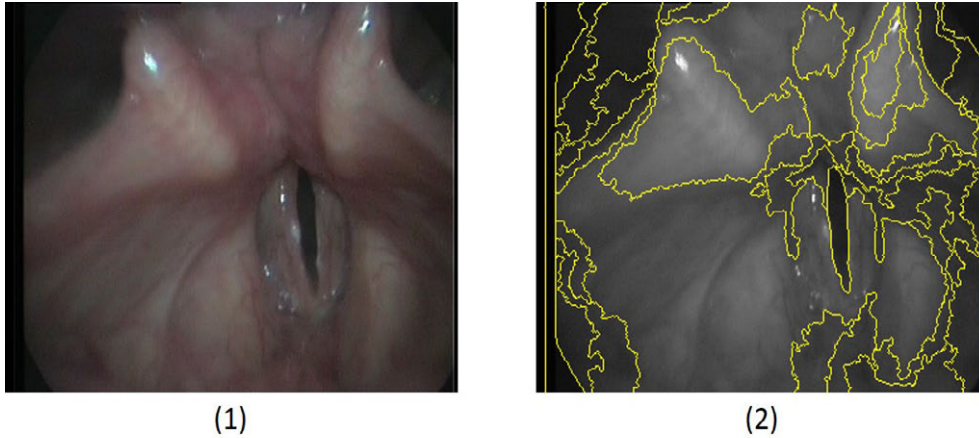


Figura 2.23: Ejemplo de merging

2.4.4. "Merging" de regiones vecinas

El siguiente paso consiste en reducir aún más el número de objetos de la imagen fusionando aquellas cuencas vecinas que posean un nivel de gris medio inferior al de la cuenca analizada en ese momento.

El objetivo de este proceso es facilitar la distinción de la glotis, ya que en principio, se espera que ésta sea un objeto con nivel de gris inferior (más oscuro) que los de su alrededor. El algoritmo es el siguiente: se toman dos cuencas y se comprueba si alguna de las dos tiene algún objeto vecino con un nivel de gris inferior. En caso de no existir, dicha cuenca es candidata a ser considerada glotis, por lo que se le atribuye un valor de la función de coste máximo, para evitar su posterior unión con otras cuencas. En caso contrario las dos cuencas analizadas quedarían unidas, con un valor de coste igual a la resta de ambos niveles de gris. Siguiendo el ejemplo anterior, en la figura 2.24 se puede ver la imagen una vez procesada con este método:

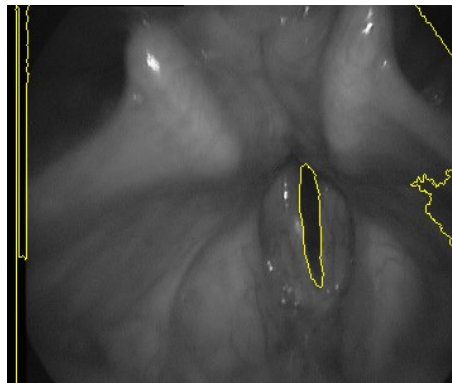


Figura 2.24: Merging de regiones vecinas

2.4.5. Predictor lineal

Esta fase tiene como objetivo distinguir la glotis del resto de elementos de la imagen por su forma. El proceso consiste en ir eliminando objetos hasta que quede únicamente el objeto glotis. Por ejemplo, el fondo se elimina porque el tamaño de área es mucho más grande que la del resto de objetos. Los laterales se eliminan por su nivel de gris, que suele ser inferior al nivel de gris que tiene la glotis (30 en las imágenes más oscuras y 70 en las más claras).

Además de los elementos mencionados, en la imagen a procesar, existen otros dos objetos: las sombras y la glotis. Para distinguir el elemento glotis entre ellos se hace uso de un clasificador, que devolverá una función de Fisher con un histograma en el que se presentan los diferentes objetos que pueden ser clasificados como glotis. Para asegurar que

ninguna glotis queda excluida en este proceso, se ha considerado la posibilidad de clasificar como glotis objetos sombra por encima de la posibilidad contraria. Una vez obtenido el histograma, si aún no se hubiese detectado la glotis, se establece que será el objeto glotis aquel que, teniendo 3 posibilidades más de ser glotis que el resto de elementos, tenga más profundidad. Entendiendo por profundidad la diferencia media entre los valores del gradiente del borde del objeto y el valor mínimo en el interior del mismo.

Con este método, se consigue un porcentaje de acierto en la detección del objeto glotis de un 98,98 % de las 110 imágenes estudiadas.

Capítulo 3

Descripción de la solución propuesta

3.1. Introducción

El conjunto de la solución se va a dividir en dos fases claramente diferenciadas. En la primera se explicará el procedimiento llevado a cabo para completar la migración de la aplicación inicial. Y en segundo lugar se comentarán en detalle las mejoras introducidas para lograr una óptima detección de la glotis.

3.2. Migración de código

Esta etapa del proyecto se hizo conjunta con el alumno Stepan Venzik. Por lo que no se van a explicar todos y cada uno de los bloques que componen la aplicación.

La aplicación de partida estaba compilada en el entorno de Borland¹ y el lenguaje de programación utilizado es C++ [18]. Dicha aplicación se migrará a la plataforma *Visual Studio 2015*, en lenguaje C#, ya que es un entorno multiplataforma ampliamente utilizado y se dispone de experiencia previa.

La principal diferencia entre estos dos lenguajes de programación es que en C++ cada clase está formada por dos ficheros - *fichero1.h* y *fichero2.cpp* - en el primero de ellos se declaran todas y cada una de las variables y funciones que posteriormente se desarrollarán en el segundo. En cambio, C# es un lenguaje de programación orientado a objetos, diseñado para compilar aplicaciones diseñadas en .NET, como es el caso [15]. En este

¹Borland Software Corporation es una compañía de software fundada en 1983.

tipo de programación aparecen nuevos conceptos y términos como objetos, propiedades, métodos, etcétera [12]. Por lo que se hace necesario explicar los pasos necesarios que se han llevado a cabo para lograr la migración de la aplicación completa.

Otro de los cambios que se ha realizado a nivel global es eliminar la definición de *fastcall* de las funciones. El aplicar esta llamada a una función determinada, significa que los primeros 3 registros de dicha llamada se almacenan en registros - específicamente en EAX, EDX y ECX respectivamente - para así hacer las llamadas mucho más rápido [14]. Esto en *C#* no es necesario, por lo que se han eliminado.

Este bloque se ha dividido en 3 subapartados con el fin de explicar mejor la migración de cada una de las partes de la aplicación. A continuación se explicará más en detalle cada uno de ellos.

3.2.1. Bloque imagen

Este primer bloque lo compone la clase *TImagen*. En esta clase se construye la imagen con la que se trabajará posteriormente. Está compuesta de distintos parámetros: estructuras, clases, variables, etc. con los que se consigue definir por completo la imagen a analizar.

Las imágenes pueden leerse desde diferentes fuentes. Por norma general, la aplicación está preparada para que el usuario elija un archivo *.bmp*, es decir, una imagen en escala RGB, o lo que es lo mismo, una matriz de 3 dimensiones. No obstante también se puede partir de una imagen de *Matlab*, siendo esta un archivo de tipo *.img* definido en escala de grises - matriz de dos dimensiones-. En función del tipo de archivo de partida, se considerará a la imagen de forma diferente, pudiendo ser de tipo *vi2D* para una imagen en escala de grises y *vi3D* para una imagen en escala de color ².

Como se ha comentado, la aplicación trabaja continuamente con matrices. Pudiendo ser estas de dos o tres dimensiones. Es uno de los inconvenientes que han surgido a la hora de migrar el código, ya que este tipo de datos se define diferente para cada lenguaje de programación. Por ejemplo, una matriz de 3 dimensiones de datos enteros en *C++* se escribiría de la forma: *int *** Imagen3D* mientras que en *C#* se pueden definir de dos maneras distintas: *int[,] Imagen3D* o *int[] [] [] Imagen3D*. La principal diferencia es que con la primera forma hay que trabajar con la matriz completa, mientras que si se utiliza la segunda, se puede trabajar con una *sub matriz* de una o dos dimensiones en función de las necesidades del momento.

Debido a este cambio en la nomenclatura de las matrices, y por consiguiente de su

²Estructuras internas de la clase *TImagen* de 2 y 3 dimensiones respectivamente

uso, se han tenido que modificar los métodos desarrollados en esta clase.

Otro *problema* añadido al realizar la migración fue que en *C#* no existen los punteros, por lo que para llevar a cabo el correcto desarrollo de la aplicación se han sustituido por parámetros de referencia. Es decir, en lugar de una variable apuntando a una dirección de memoria en la que el método en cuestión guarde el valor devuelto, se pasa como parámetro una variable adicional de salida en la que se guarda dicho resultado.

Para el manejo de imágenes, la aplicación inicial hace uso una segunda clase llamada *Funciones*, en la que están codificadas distintas funciones: Inicializar matrices, copiar datos entre matrices, distintas operaciones con matrices... La migración de esta clase no ha tenido más problemas que los comentados hasta ahora.

3.2.2. Bloque reserva

Este bloque lo componen a su vez tres clases diferentes, *Reserva1D*, *Reserva2D* y *Reserva3D*. Cada una de ellas está implementada para reservar memoria para estructuras de una, dos o tres dimensiones respectivamente. Están preparadas para admitir cualquier tipo de datos estándar.

Cada una de ellas consta de distintos métodos para operar con las estructuras - matrices - mencionadas - como puede ser redimensionarlas - así como constructores y destructores.

Otra de las grandes diferencias entre los dos lenguajes de programación mencionados, es que en *C++* se trabaja con manejo de memoria - en la aplicación inicial se utilizan funciones como *calloc* o *realloc* - mientras que en *C#* no es necesario, ya que se reserva memoria implícitamente al declarar cada una de las variables (estructuras) con la instancia *new*. Por esta razón, en la migración de cada una de las 3 clases comentadas, se han descartado los métodos destructores.

Otro detalle es que tampoco se han desarrollado los métodos *Asigna* programados en la aplicación inicial, ya que no eran necesarios.

Por lo tanto, la nueva versión de cada una de las tres clases *Reserva* queda formada por: un constructor - cada uno para una determinada dimensión - compuesto por los datos de la estructura y las dimensiones de la misma, un método llamado *Redimensionar* que, como su propio nombre indica se utiliza para redimensionar una determinada estructura, y por último otro método llamado *DevuelvePuntero*, que devuelve los datos de la estructura.

3.2.3. Bloque inicio

Este bloque se compone de la ventana principal de la aplicación. Será el medio de interacción entre el usuario y la aplicación, donde este elegirá los parámetros que desee para posteriormente realizar el análisis de cada una de las imágenes.

Al migrar esta parte de la aplicación, no han surgido mayores complicaciones que la diferencia entre las nomenclaturas de los controles para cada uno de los dos lenguajes de programación.

Este bloque se explica más en detalle en el apartado 5.1.

3.3. Técnicas de preprocesado

Para probar cada una de estas mejoras primero se realizó un estudio de las imágenes obtenidas mediante *scripts* de *Matlab*. Y una vez comprobados los resultados, se procedió a hacer la migración a la aplicación final de aquellas en las que se obtuvo mejor resultado.

3.3.1. Normalización Localizada (LN)

Se trata de un algoritmo de compensación de iluminación. Este algoritmo permite recuperar información de las zonas más oscuras de la imagen, por lo que puede ser interesante a la hora de destacar el objeto glotis sobre la imagen para su posterior procesamiento [8].

Los pasos que se han realizado para aplicar el algoritmo son:

1. Definir un ancho de ventana
2. Calcular el promedio en la vecindad del píxel definida por la ventana centrada en (x,y)
3. Calcular la desviación estándar en la vecindad del píxel definida por la ventana centrada en (x,y)
4. Calcular el valor de intensidad compensada de la imagen

Este parámetro se calcula con la siguiente expresión:

$$I_{LN}(x, y) = \frac{I(x, y) - E(I(u, v))}{\sigma(I(u, v))}(u, v) \in W \quad (3.1)$$

5. Aplicar estiramiento lineal para regresar a la escala de grises.

Preprocesado para la mejora de la segmentación del área glotal.

En la figura 3.1 se muestran los resultados obtenidos al aplicar este algoritmo sobre una imagen con distintos tamaños de ventana:

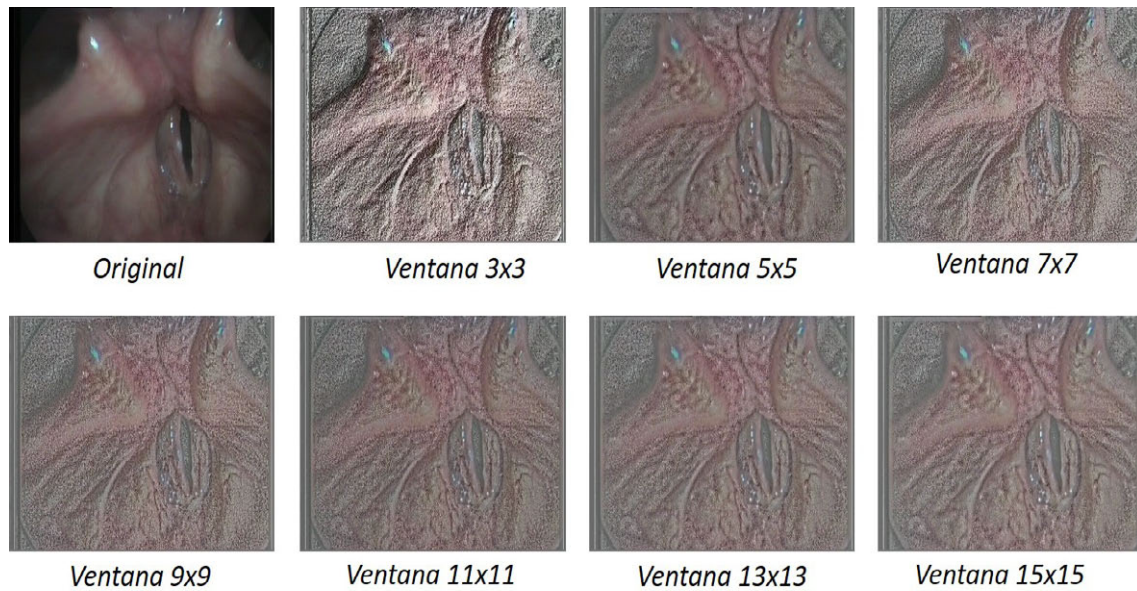


Figura 3.1: Aplicación del algoritmo LN para diferentes tamaños de ventana

Observando los resultados obtenidos, se puede ver que probablemente no sirva para el análisis requerido, ya que el objeto glotis, a simple vista, toma valores de niveles de color parecidos al resto de objetos de la imagen, por lo que se hace más difícil detectarlo.

3.3.2. Transformada Discreta del Coseno

La *Transformada Discreta del Coseno*, descubierta por Ahmed et al.- de ahora en adelante DCT ³ - es una técnica basada en la *Transformada Discreta de Fourier* pero utilizando números reales. La descomposición en factores se hace exclusivamente con sumas de cosenos, no de senos y cosenos [7][8].

El objetivo de utilizar la DCT es realizar la compensación de iluminación mediante la anulación de las componentes de baja frecuencia en el dominio transformado.

Pasos realizados:

1. Convertir imagen RGB a imagen YCbCr:

Para poder aplicar la DCT hay que transformar a YCbCr. Al convertir las imágenes, obtenemos por un lado la componente de luminancia - Y -, y por otro lado las componentes de crominancia - Cb y Cr - que son la diferencia de nivel de azul y de nivel de rojo respectivamente.

Para realizar la conversión se han utilizado las siguientes expresiones [17]

$$Y = 16 + (0,257 \cdot R) + (0,504 \cdot G) + (0,098 \cdot B) \quad (3.2)$$

$$Cb = 128 + (0,148 \cdot R) - (0,291 \cdot G) + (0,439 \cdot B) \quad (3.3)$$

$$Cr = 128 + (0,439 \cdot R) - (0,368 \cdot G) - (0,071 \cdot B) \quad (3.4)$$

2. DCT

Para obtener la DCT se han aplicado las siguientes fórmulas sobre la matriz de entrada, obteniendo una matriz de iguales dimensiones:

$$B_{pq} = \alpha_p \cdot \alpha_q \cdot \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos\left(\frac{\pi(2m+1)p}{2M}\right) \cos\left(\frac{\pi(2n+1)q}{2N}\right) \quad (3.5)$$

$$Tomando : \begin{cases} 0 \leq p \leq (M-1) \\ 0 \leq q \leq (N-1) \end{cases} \quad (3.6)$$

³De las siglas en inglés Discrete Cosine Transform

$$\alpha_p \begin{cases} \frac{1}{\sqrt{M}} & \text{si } p = 0 \\ \sqrt{\frac{2}{M}} & \text{resto} \end{cases} \quad (3.7)$$

$$\alpha_q \begin{cases} \frac{1}{\sqrt{N}} & \text{si } q = 0 \\ \sqrt{\frac{2}{N}} & \text{resto} \end{cases} \quad (3.8)$$

3. Multiplicación por la matriz triángulo

Se define una matriz de iguales dimensiones que la matriz a transformar, y en la esquina superior izquierda, se crea un triángulo de ceros de lado predefinido a excepción del primer parámetro. Por ejemplo, para una matriz de 5x5, definimos un triángulo de lado 3 y queda:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (3.9)$$

Al multiplicar la matriz triángulo por la matriz resultado de la DCT, se consigue anular los coeficientes de baja frecuencia de la imagen transformada que caen dentro del triángulo.

4. IDCT

Se reconstruye la matriz inicial haciendo la inversa de la transformada discreta del coseno.

5. Convertir imagen YCbCr a imagen RGB

Se convierte la matriz resultado de la IDCT, cuyos niveles de color están expresados en escala YCbCr como se ha comentado. Para eso se tiene que tener en cuenta que la escala RGB toma valores entre 0 y 255, por lo tanto, las fórmulas tienen que devolver siempre un valor dentro de ese rango. Se utilizarán las siguiente ecuaciones, tomando Y' , Cb' , y Cr' como $(Y-16)$, $(Cb-128)$ y $(Cr-128)$ respectivamente [17]:

$$R = \text{Maximo}(((1,164 \cdot Y') + (0,000 \cdot Cb') + (1,596 \cdot Cr')), 0) \quad (3.10)$$

$$G = \text{Maximo}(((1,164 \cdot Y') - (0,392 \cdot Cb') - (0,813 \cdot Cr')), 0) \quad (3.11)$$

$$B = \text{Maximo}(((1,164 \cdot Y') + (2,017 \cdot Cb') + (0,000 \cdot Cr')), 0) \quad (3.12)$$

Una vez implementada la función, se han hecho pruebas para distintas longitudes de lado de triángulo: 3, 6, 9, 12, 15, 18, 21.

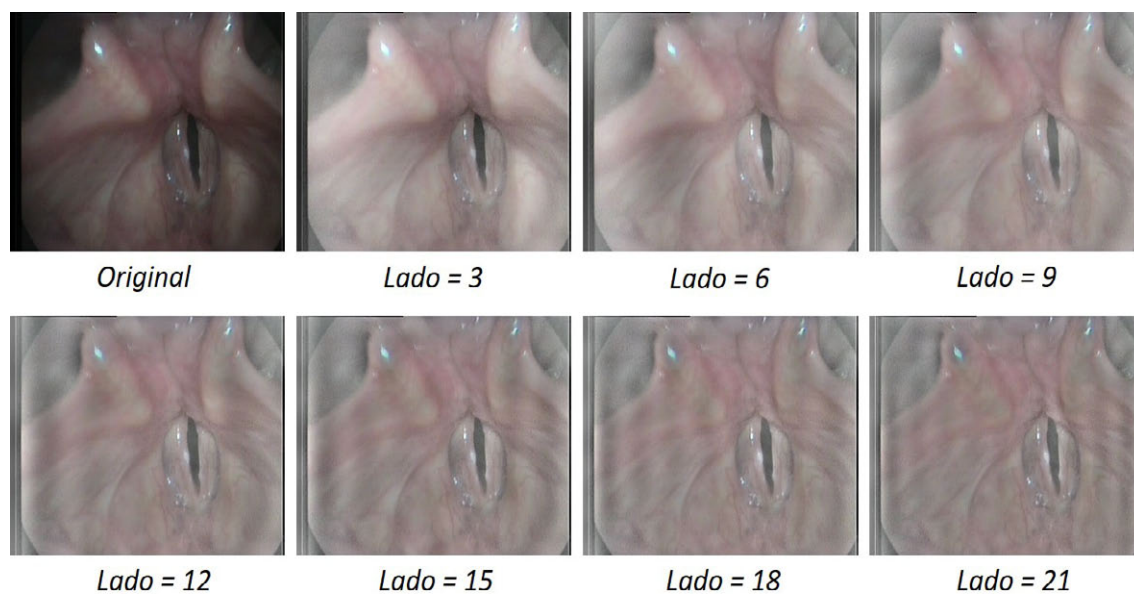


Figura 3.2: Aplicación del algoritmo de la DCT para distintos tamaños del lado del triángulo

De los resultados obtenidos cabe destacar que a medida que se aumenta el tamaño del lado del triángulo, las imágenes presentan zonas con oscilaciones, lo que puede implicar pérdida de información importante. No obstante, para el análisis que se va a realizar en este proyecto, no se puede descartar ningún resultado, ya que se busca que el objeto glotis de la imagen tenga niveles de color claramente diferenciados del resto de objetos de la imagen.

3.3.3. Filtro anisotrópico

El filtrado anisotrópico consiste en mejorar la calidad de una imagen que está vista desde un ángulo oblicuo con respecto al ángulo de proyección de la imagen sobre una superficie. Este tipo de filtro elimina el efecto “aliasing”[19].

Este algoritmo se desarrolló en el proyecto fin de grado realizado por la alumna: María Cristina Marco Llorente. En este proyecto se implementó el filtrado mediante un *script* de *Matlab*, ahora se ha migrado al entorno actual con el fin de incluirlo en la aplicación.

El algoritmo necesario para llevar a cabo este tipo de filtrado es muy pesado. Consiste en modificar el valor de cada pixel de la imagen por otro dado por una función que varía dependiendo del valor del nivel de color del propio pixel y el de sus vecinos más cercanos en horizontal y en vertical.

La nomenclatura a utilizar en el desarrollo de este algoritmo se muestra en la siguiente imagen, siendo I el nivel de color de cada píxel. Para la nomenclatura de los píxeles vecinos, se han utilizado los cuatro puntos cardinales (Norte - Superior, Sur - Inferior, Este - Derecha y Oeste Izquierda). Así mismo, se ha denominado C al parámetro que relaciona cada pixel vecino con el central.

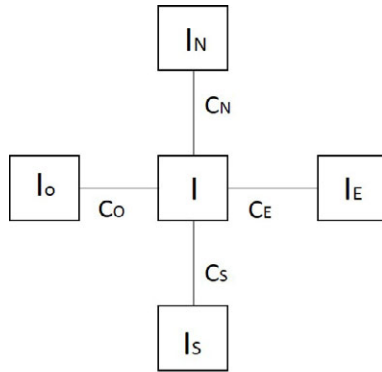


Figura 3.3: Nomenclatura utilizada para el filtrado anisotrópico.

Para el cálculo del nuevo valor de I del pixel analizado - ID - se utiliza la siguiente expresión:

$$I_D = I + \lambda \cdot [C_N \cdot D_N + C_S \cdot D_S + C_E \cdot D_E + C_O \cdot D_O] \quad (3.13)$$

Siendo λ un parámetro de control de difusión que varía entre 0 y 0.25, y D la diferencia de nivel de color entre el píxel central y cada píxel vecino:

$$D_i = I_i - I \quad (3.14)$$

El valor de C se calcula con la siguiente expresión, siendo K una constante:

$$C_i = e^{(-\frac{|I_i|}{K})^2} \quad (3.15)$$

El algoritmo utilizado para realizar el filtrado es largo y costoso, pues hay que recorrer la matriz un determinado número de veces para cada una de las 3 capas RGB. Cada vez que se recorre la matriz, se sustituyen los valores de cada píxel por el calculado utilizando las ecuaciones anteriores.

Para una mejor comprensión, se muestra el *pseudocódigo* implementado para realizar este algoritmo:

FOR capas RGB

FOR numero de repeticiones

Se recorre la matriz inicial y se reasignan valores

END FOR

END FOR

A continuación se muestran unas imágenes obtenidas mediante este método. Se han realizado 5 pruebas, cada una de ellas con distintos parámetros. Los parámetros que han variado son: K y N , que es el número de repeticiones anteriormente mencionado. Mientras que λ se ha mantenido constante con un valor de 0.05.

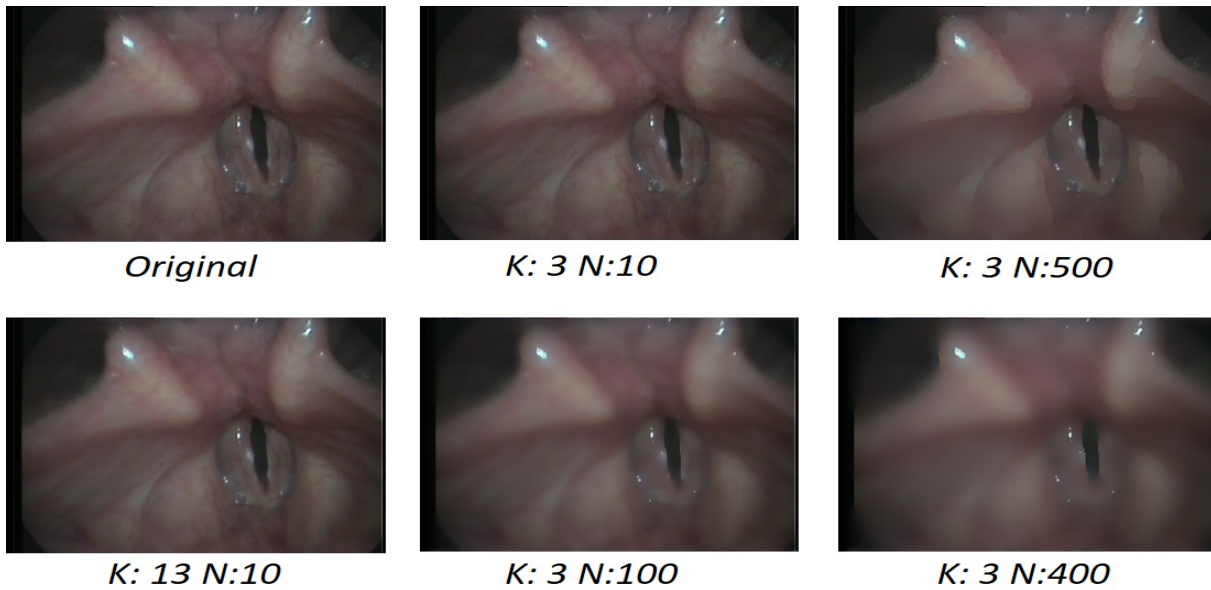


Figura 3.4: Ejemplos de imágenes aplicándoles el filtrado anisotrópico con distintos parámetros.

Como se puede observar en las imágenes anteriores, a medida que se aumenta el número de repeticiones, las imágenes quedan mucho más difuminadas perdiendo nitidez. Pero este tampoco es un resultado descartable, ya que, como se ha indicado en otros apartados, para este estudio interesa que la glotis se diferencie el máximo posible del resto de objetos de la imagen.

Capítulo 4

Resultados

4.1. Introducción

Una vez se ha analizado en detalle cada una de las mejoras aportadas a la aplicación inicial, se va a proceder a explicar los resultados obtenidos tanto con cada una de ellas por separado como con todas ellas en conjunto.

El capítulo se va a dividir en dos apartados, en el primero de ellos se van a mostrar los resultados obtenidos al migrar el código de la aplicación inicial, y en el segundo, se analizarán los resultados obtenidos al aplicar cada una de las mejoras implementadas.

Para el análisis de resultados se ha partido de una biblioteca de imágenes de distintos vídeos estrobolaringscópicas utilizadas en la tesis de partida - 110 imágenes extraídas de 15 vídeos -, en las que varía la luminosidad de la imagen así como la nitidez de la misma. Puede que para algunas de ellas sea mejor utilizar un tipo de preprocesado y para otras otro en función de sus características.

4.2. Resultados de la migración del código

Para comprobar el correcto funcionamiento de la aplicación migrada se hará uso de la base de datos mencionada. En el procesado de cada una de las imágenes se aplican las dos fases del “Merging” explicadas en los apartados [2.4.4] y [2.4.3] respectivamente. Tal y como se explica en el apartado [3.2], las modificaciones introducidas se traducen en segmentaciones ligeramente diferentes.

En el proceso de comparación de los resultados se aprecia que se obtienen mejores efectos en la detección de la glotis en 15 imágenes de las 110 totales (13.63 %). Sin embargo, existen errores debidos a causas ajenas a la aplicación, como pueden ser: la mala iluminación - que introduce sombras, que unidas con la glotis, dificultan su correcta detección - o la división de la glotis.

Por otro lado, en 10 imágenes del total (9.09 %), se aprecia una mejor detección del área glotal.

En las siguientes figuras se pueden ver algunos ejemplos de las segmentaciones realizadas con las dos aplicaciones. Las imágenes de la fila superior - líneas pintadas de azul - se corresponden a análisis realizados con la aplicación original, mientras que las que están situadas en la fila inferior - líneas amarillas - han sido procesadas mediante la aplicación migrada.

En la figura [4.1] se pueden apreciar algunos ejemplos de segmentaciones en las que no se ha conseguido alcanzar el resultado de la aplicación inicial, mientras que en la [4.2] se muestran otros ejemplos en los que se ha mejorado.

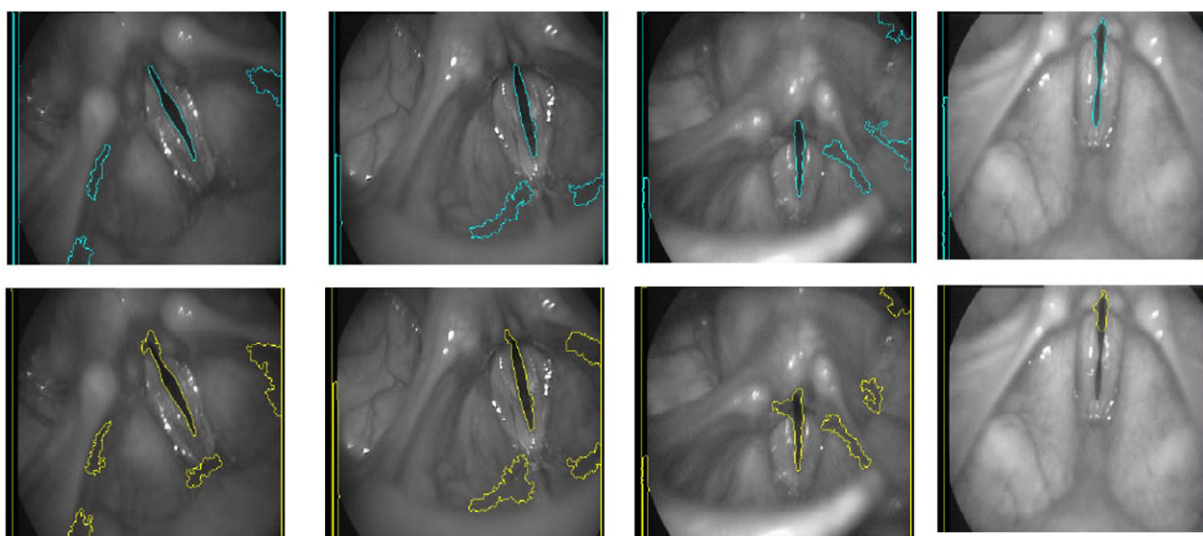


Figura 4.1: Resultados migración con errores

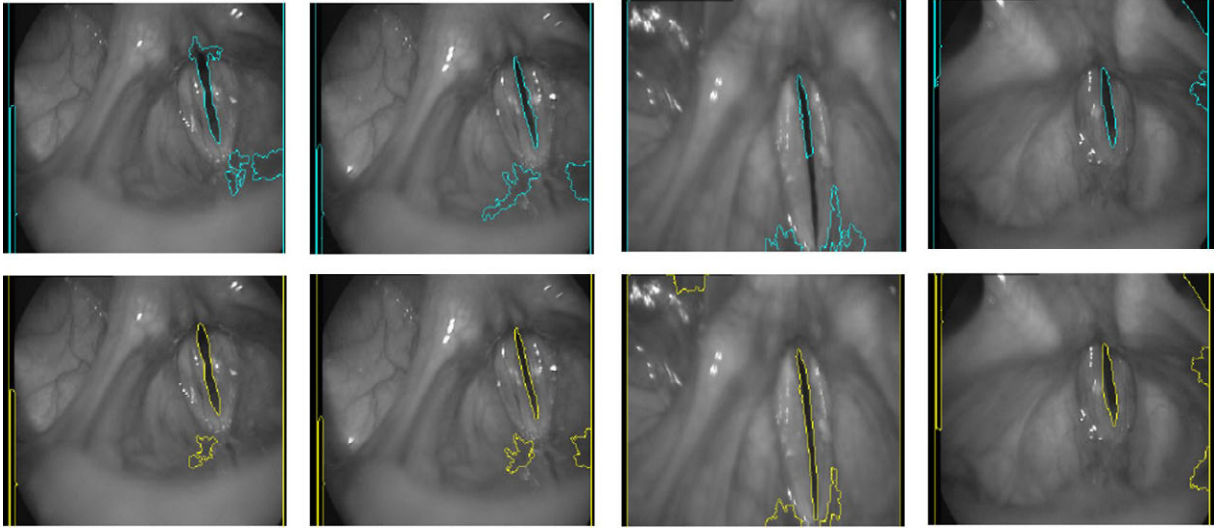


Figura 4.2: Resultados migración con mejoras

4.3. Resultados de las mejoras implementadas

De nuevo, para el análisis de cada una de las mejoras implementadas, se hará uso de la base de datos mencionada anteriormente. Se debe tener en cuenta que no todas ellas cuentan con la misma iluminación ni la misma nitidez. Por eso para algunas de ellas será mejor utilizar un tipo de preprocesado y para otras otro en función de sus características.

Los resultados con los que se va a comparar si existe verdaderamente una mejora son los obtenidos mediante la misma aplicación introduciéndole un análisis con información de color.¹

Para corroborar las mejoras implementadas, se va a utilizar un *script* programado en *Matlab*. Este *script* analiza cada una de las imágenes obtenidas, comparándolas con las mismas imágenes simulando resultados ideales - imágenes segmentadas por un experto -, y devolviendo una serie de parámetros que finalmente dirán si las mejoras efectivamente introducen mejoras en la detección del área glotal, como pueden ser: *Jaccard index*, *DICE coeff*², precisión, sensibilidad...

¹Resultados de la aplicación de Stefan Venzik, manteniendo los parámetros por defecto

²Coficiente o índice de Sorensen-Dice, utilizado para medir la similitud entre dos muestras. Muy parecido al Jaccard index.

De los parámetros que devuelve dicho *script*, se va a analizar el *Jaccard index*. Este parámetro mide el grado de similitud entre dos conjuntos - A y B - [16]. El resultado es la intersección de ambos conjuntos dividida entre la su unión:

$$J(A, B) = |A \cap B| / |A \cup B| \quad (4.1)$$

El índice de Jaccard toma valores comprendidos entre 0 y 1, correspondiéndose el valor cero a que no tienen nada en común y el valor uno a que son idénticos.

En este apartado se van a analizar por separado cada una de las mejoras, para posteriormente hacer un análisis de aquellas que han introducido mejoras en conjunto. Cabe destacar que para los primeros análisis - [4.3.1], [4.3.2], [4.3.3] y [4.3.4] - se han mantenido constantes los parámetros configurables de la aplicación, para así comprobar también el grado de autonomía de la misma. Por último se expondrán algunos ejemplos de análisis modificando estos parámetros [4.3.5].

4.3.1. Análisis de resultados para: LN

Tal y como se comentó en el apartado [3.3.1], los resultados con las imágenes obtenidas por el *script* de *Matlab* no producían una mayor diferenciación de la glotis, por lo que primero se introdujeron dichas imágenes en la aplicación para hacer un primer examen eliminatorio.

A continuación se muestran algunos ejemplos de estos análisis, en la fila superior - líneas azules - se muestran las imágenes analizadas por el software sin aplicar el algoritmo, mientras que en la fila inferior - líneas amarillas - se pueden ver las imágenes una vez aplicado:

4.3. RESULTADOS DE LAS MEJORAS IMPLEMENTADAS

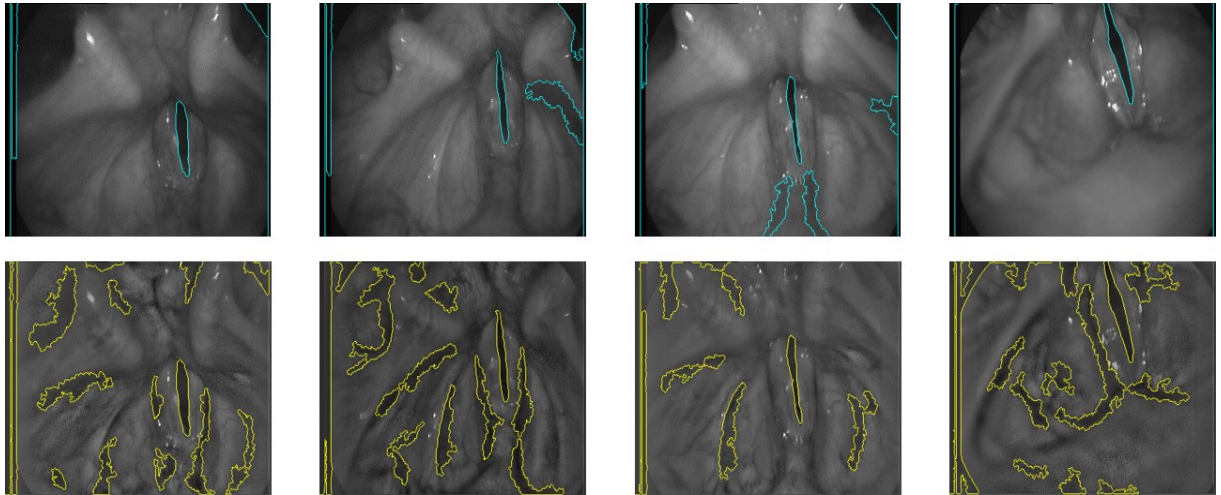


Figura 4.3: Resultados algoritmo LN: Glotis detectada

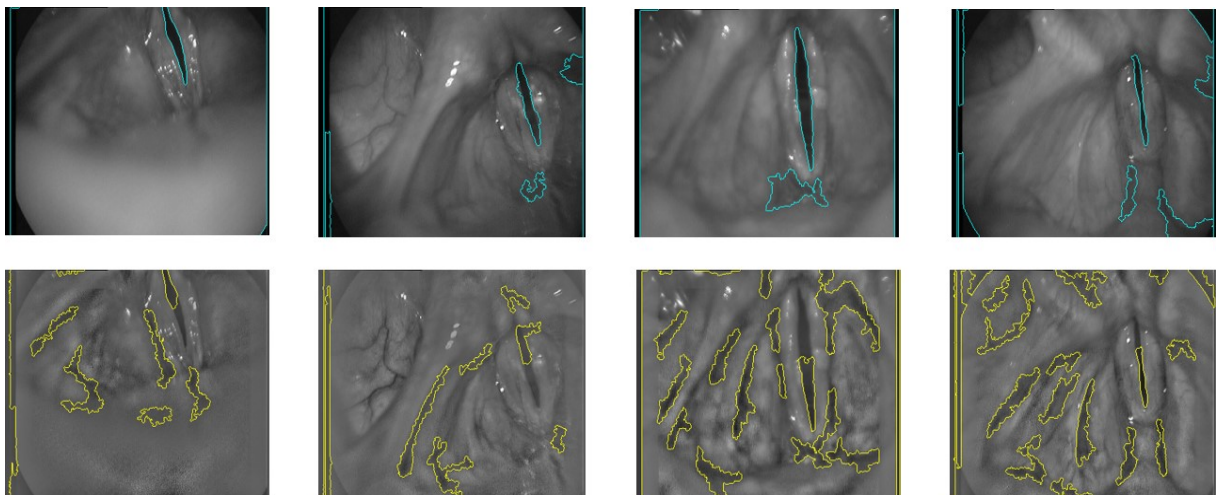


Figura 4.4: Resultados algoritmo LN: Glotis no detectada

Como se puede ver en la figura [4.3], mediante este algoritmo se consigue detectar la glotis en la imagen analizada en muchas imágenes, pero introduce diferentes elementos que dificultan el etiquetado de la misma como el *objeto glotis*. Por el contrario, en la figura [4.4], se puede ver otros ejemplos en las que la aplicación no es capaz de detectarla.

Después de realizar varios análisis con distintos tamaños de ventana, se descartó este algoritmo, puesto que no ofrece mejoras destacables respecto de la aplicación inicial. Se detecta la glotis únicamente en 7 de las 110 imágenes disponibles (6.36 %)

4.3.2. Análisis de resultados para: DCT

En esta sección se explicarán los resultados obtenidos aplicando el algoritmo explicado en el apartado [3.3.2]. Como ya se comentó, este algoritmo proporciona una mejor uniformidad en el color de la imagen inicial.

Para el análisis de resultados con este algoritmo, se han hecho pruebas con distintos valores del tamaño del lado del triángulo. En la siguiente tabla se muestran los valores obtenidos.

MEDIAS	Jaccard index				
	ORIGINAL	Tr=9	Tr=12	Tr=15	Tr=18
V0	0,768	0,825	0,876	0,801	0,750
V1	0,773	0,767	0,708	0,850	0,843
V2	0,730	0,587	0,594	0,610	0,496
V3	0,812	0,623	0,826	0,812	0,816
V4	0,120	0,488	0,629	0,702	0,636
V5	0,868	0,603	0,763	0,545	0,743
V6	0,895	0,864	0,863	0,847	0,659
V7	0,706	0,584	0,502	0,731	0,576
V8	0,433	0,445	0,349	0,521	0,620
V9	0,484	0,795	0,781	0,777	0,778
V10	0,760	0,502	0,637	0,635	0,590
V11	0,774	0,776	0,780	0,739	0,801
V12	0,607	0,736	0,581	0,336	0,622
V13	0,302	0,693	0,481	0,780	0,640
V14	0,426	0,579	0,816	0,747	0,692
TOTAL	0,631	0,658	0,679	0,696	0,684

Figura 4.5: Tabla de resultados DCT

En la tabla 4.5 se han representado las medias de los resultados obtenidos para el parámetro *Jaccard index* para las imágenes de cada vídeo. En la primera columna aparece el nombre del vídeo, la siguiente se corresponde con el valor obtenido para las imágenes originales y por último, en las demás columnas, se reflejan los valores obtenidos para el análisis realizado con distintos tamaños del lado del triángulo - 9, 12, 15 y 18 -.

4.3. RESULTADOS DE LAS MEJORAS IMPLEMENTADAS

Atendiendo a los datos mostrados en la tabla se puede ver que los resultados mejoran en 11 vídeos de los 15 totales (73,33 %), siendo los mejores resultados, los obtenidos aplicando un lado del triángulo 12 o 15. Este último es el que mejor media presenta. No obstante, en el anexo se puede ver una tabla (7.1) en la que quedan reflejados los datos obtenidos para cada una de las imágenes.

En la figura 4.6 se muestran algunos ejemplos de segmentaciones aplicando un triángulo de lado 15, que es para el que mejores resultados se han obtenido. De nuevo en la fila superior se sitúan las imágenes obtenidas sin preprocesado, y en la fila inferior las imágenes aplicándoles la transformada.

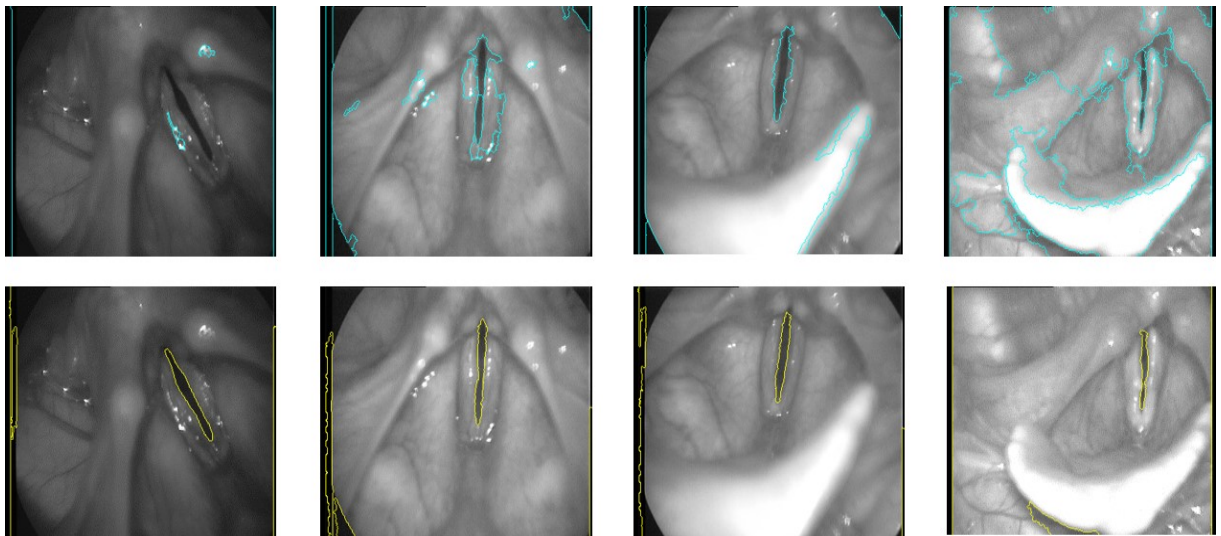


Figura 4.6: Resultados DCT

En el análisis global, cabe destacar que aplicando este preprocesado, se consigue reducir a 10 el número de imágenes en las que no se detecta la glotis en el caso del triángulo de lado 15 y a 15 en el caso del triángulo de lado 12, siendo 22 el número de imágenes no detectadas sin aplicar preprocesado.

4.3.3. Análisis de resultados para: Filtrado anisotrópico

Siguiendo el mismo esquema que en el apartado anterior, ahora se van a exponer los resultados obtenidos aplicando un preprocesado basado en un filtrado anisotrópico, explicado en el apartado [3.3.3].

Nuevamente se han realizado pruebas para distintos valores tanto del número de repeticiones como del coeficiente de difusión - N y K respectivamente -. En la siguiente tabla se pueden ver los resultados obtenidos para las medias de los vídeos.

MEDIAS	Jaccard index					
	ORIGINAL	K=3 N=10	K=3 N=500	K=13 N=10	K=13 N=100	K=13 N=400
V0	0,768	0,638	0,483	0,385	0,000	0,000
V1	0,773	0,774	0,863	0,444	0,000	0,208
V2	0,730	0,604	0,707	0,630	0,126	0,126
V3	0,812	0,641	0,846	0,877	0,597	0,401
V4	0,120	0,477	0,412	0,413	0,450	0,184
V5	0,868	0,876	0,838	0,879	0,000	0,000
V6	0,895	0,871	0,869	0,895	0,634	0,248
V7	0,706	0,734	0,695	0,670	0,000	0,000
V8	0,433	0,437	0,659	0,817	0,344	0,121
V9	0,484	0,582	0,281	0,801	0,213	0,000
V10	0,760	0,715	0,476	0,753	0,243	0,211
V11	0,774	0,781	0,815	0,682	0,000	0,000
V12	0,607	0,736	0,467	0,620	0,128	0,000
V13	0,302	0,336	0,631	0,751	0,735	0,525
V14	0,426	0,551	0,576	0,452	0,123	0,124
TOTAL	0,631	0,650	0,641	0,671	0,239	0,143

Figura 4.7: Tabla de resultados filtrado anisotrópico

En la tabla 4.7 están representados los valores obtenidos para el parámetro *Jaccard index* para las medias de las imágenes de cada vídeo. Se ha seguido la misma nomenclatura que en el apartado anterior, siendo las dos primeras columnas los valores de los nombres de cada vídeo y las medias de las imágenes originales, y las siguientes columnas los valores aplicando diferentes parámetros.

4.3. RESULTADOS DE LAS MEJORAS IMPLEMENTADAS

En este caso los resultados muestran una mejora en las medias para 12 de los 15 vídeos iniciales (80 %), siendo los valores de $K=3$ $N=10$ y de $K=13$ $N=10$ los que mejores resultados obtienen. No obstante, este preprocesado afecta mucho al rendimiento de la aplicación, ya que aumenta en gran medida el tiempo de análisis de cada imagen.

En la figura 4.3.3 se muestran algunos ejemplos de este análisis para 13 repeticiones y coeficiente de difusión igual a 10:

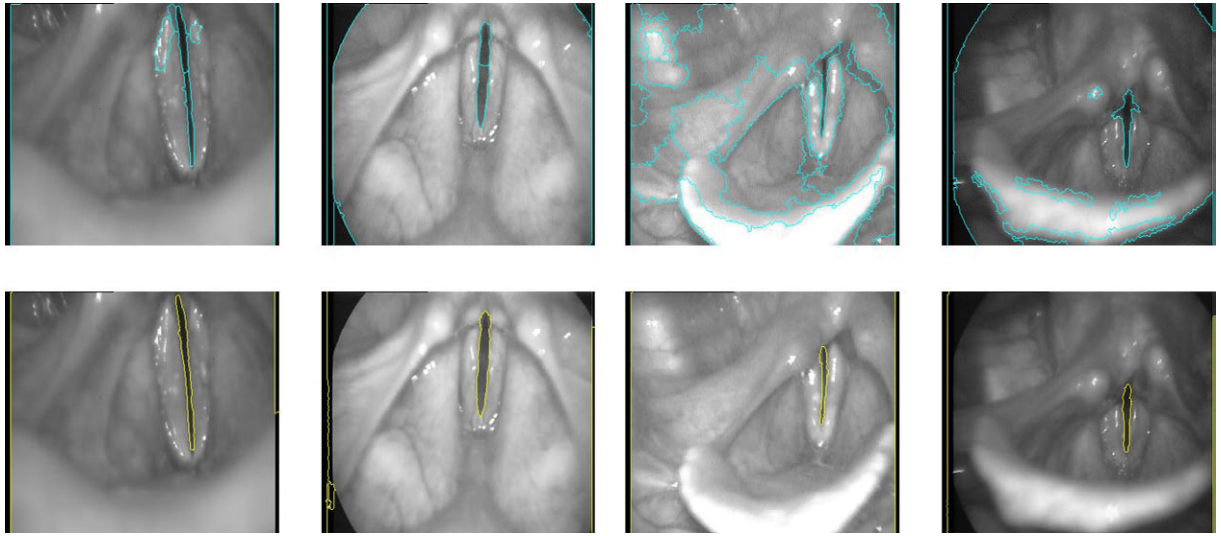


Figura 4.8: Resultados filtro anisotrópico

Analizando cada una de las imágenes, se puede ver que el número de imágenes en las que se detecta la glotis es similar aplicando o no el preprocesado, por lo que en este aspecto no presenta una mejoría notable. Lo que si cabe destacar es que el parámetro *Jaccard index* en gran parte de las imágenes es superior al aplicar el preprocesado. En el anexo se muestra una tabla con los resultados completos (7.2)

4.3.4. Análisis de resultados conjuntos

Una vez analizado cada preprocesado por separado, se va a analizar en conjunto. Los parámetros que se han escogido para este análisis son: para el filtro anisotrópico: 13 repeticiones y coeficiente de difusión igual a 10, y para la transformada discreta del coseno un lado del triángulo igual a 12.

Los resultados se han recogido en la tabla 4.9. En este caso, para una mejor visualización, se han agregado los datos de las dos mejoras por separado en la segunda y tercera columna y la mejora conjunta en la última:

MEDIAS	ORIGINAL	DCT	FILTRO	DCT + FILTRO
V0	0,768	0,876	0,385	0,511
V1	0,773	0,708	0,444	0,858
V2	0,730	0,594	0,630	0,840
V3	0,812	0,826	0,873	0,826
V4	0,120	0,629	0,572	0,667
V5	0,868	0,763	0,879	0,862
V6	0,895	0,863	0,895	0,846
V7	0,706	0,502	0,670	0,766
V8	0,433	0,349	0,819	0,792
V9	0,484	0,781	0,715	0,796
V10	0,760	0,637	0,743	0,682
V11	0,774	0,780	0,682	0,809
V12	0,607	0,581	0,639	0,833
V13	0,302	0,481	0,732	0,727
V14	0,426	0,816	0,654	0,541
TOTAL	0,631	0,679	0,689	0,757

Figura 4.9: Tabla de resultados con ambos preprocesados

Como se puede observar, la media total al aplicar los dos preprocesados aumenta, a la vez que se reduce el número de glotis no detectadas hasta 7. La tabla con los resultados al completo está adjunta en el anexo (7.3).

Viendo los resultados, se puede apreciar que en función de las características del vídeo se hace necesario un tipo de preprocesado u otro, siendo la suma de los dos procesados el que mejora la media en mayor número de vídeos, seguido del filtro anisotrópico y por último la transformada discreta del coseno.

4.3. RESULTADOS DE LAS MEJORAS IMPLEMENTADAS

En la figura 4.10 se muestran algunos ejemplos de segmentaciones aplicando este pre-procesado:

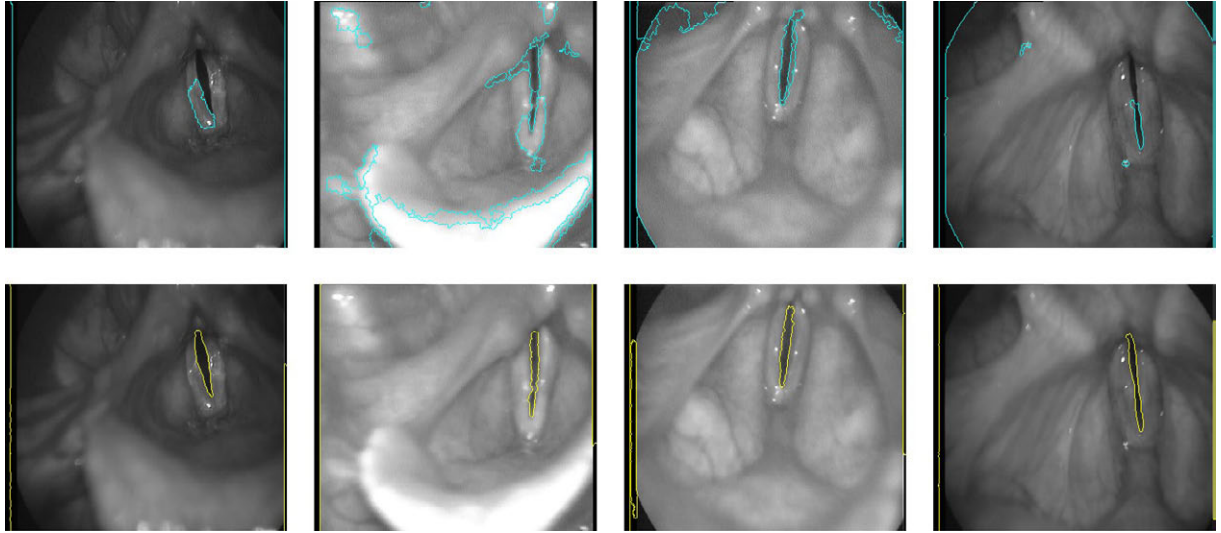


Figura 4.10: Resultados para ambos preprocesados

Para corroborar la validez de los resultados obtenidos, se hecho uso de un análisis estadístico en base a las medias de los vídeos. Se ha calculado el intervalo de confianza del conjunto de datos utilizando las fórmulas de la herramienta *Microsoft Excel* [13].

El intervalo de confianza es un intervalo de valores, en el que la media de la muestra será el centro y los valores máximo y mínimo se calculan sumándole o restándole el resultado de la función a dicha muestra.

El intervalo de confianza se ha calculado con la función *INTERVALO.CONFIANZA.NORM()* de *Excel*. Esta función tiene 3 parámetros, el primero de ellos, alfa, es el nivel de significación usado para calcular el nivel de confianza. En este caso se ha aplicado un 0.05, que indica un nivel de confianza del 95 %. El segundo parámetro es la desviación estándar - las medias de los vídeos -. Y el tercero es el tamaño de la muestra, en este caso es el número de imágenes, es decir, 110.

La fórmula del intervalo de confianza quedaría definida por la siguiente expresión:

$$\bar{x} \pm 1,96 \left(\frac{0,95}{\sqrt{110}} \right) \quad (4.2)$$

Los intervalos de confianza calculados para la media de los resultados obtenidos para las 110 imágenes han sido de 0.066 para los vídeos originales y de 0.050 para los vídeos después de aplicarles el preprocesado. En este caso, el intervalo de los vídeos originales quedaría definido entre (0.565 - 0.697), mientras que aplicándoles el preprocesado quedaría: (0.707 - 0.807). Como los dos intervalos no se solapan, se concluye que estadísticamente el análisis mejora los resultados originales.

4.3.5. Análisis de resultados modificando parámetros

Como se ha comentado en otros apartados, el análisis se ha realizado manteniendo constantes los parámetros variables de la aplicación como por ejemplo el *umbral de coste* o el *umbral del gradiente*. Esto se ha mantenido así para poder valorar la automatización de la aplicación con el preprocesado, y por eso algunos de los vídeos no mejoran tanto como cabría esperar.

En este apartado se pretende analizar los vídeos que no han obtenido los resultados esperados modificando los parámetros comentados para verificar si para dichos vídeos también es válido el preprocesado.

Se van a analizar dos vídeos, el *V0* y el *V14*, que son los vídeos que *peores* resultados tienen. En las figuras 4.11 y 4.12 se muestran los resultados obtenidos modificando el valor del umbral de gradiente a 1:

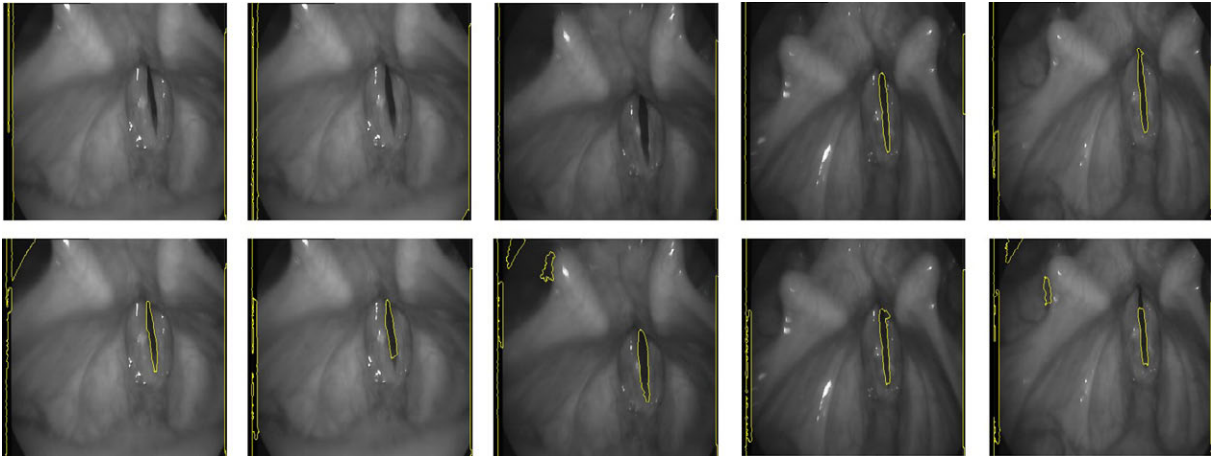


Figura 4.11: Resultados V0

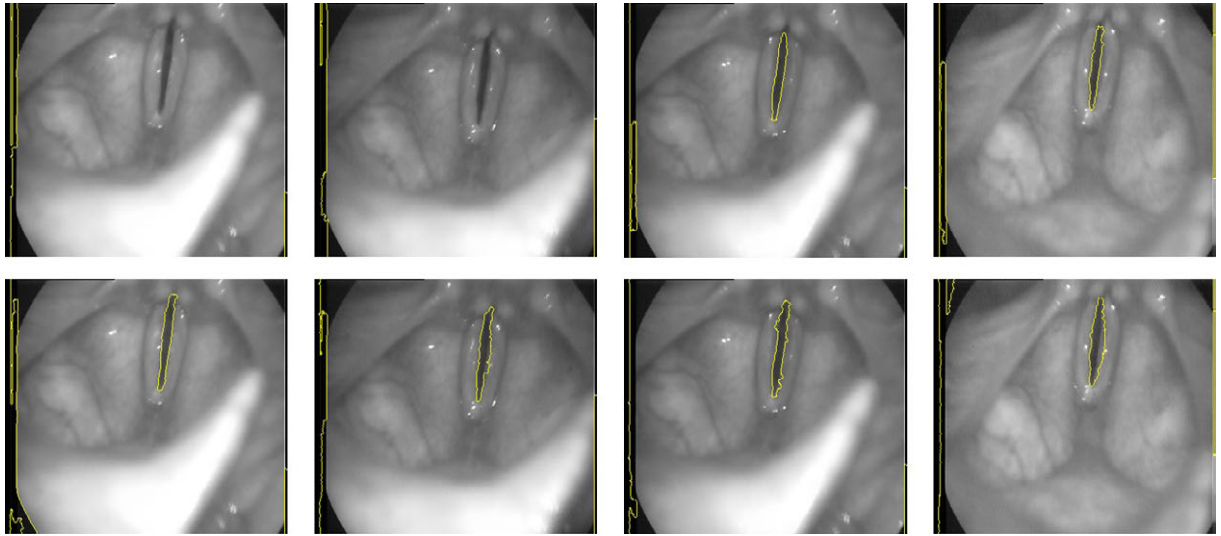


Figura 4.12: Resultados V14

Como se puede apreciar, si bien es cierto que modificando únicamente el umbral del gradiente se consiguen detectar las glotis que anteriormente habían fallado, con esta modificación empeoran los resultados obtenidos para las imágenes en las que sí que se había detectado previamente.

Capítulo 5

Manual de usuario

En esta sección se pretende dar una visión global del funcionamiento de la aplicación, indicando cada uno de los parámetros y de las funciones que se han programado. Facilitando así el trabajo de un posible usuario.

5.1. Pantalla principal

La pantalla principal se corresponde con la siguiente imagen:

Aplicación para aislamiento de glotis

Segmentación

☐ VIDEO Fichero: Explorar

☒ IMAGEN Fotogramas: ☐ Todos
(Número de fotogramas separados por comas, o un rango de fotogramas separado por un guión)

Parámetros Ajustables para Segmentación

Umbral de Gradiente:

Porcentaje de área límite:

Nivel de insensibilidad total:

Mínimo insensibilidad total:

Umbral de coste para Merging:

☒ Tomar nivel de insensibilidad total en porcentaje.

☒ Aplicar Merging huérfanas de gris al finalizar.

Información de color

☐ Aplicar Merging Color

☒ Gradiente 3D

Preprocesado

☒ Transformada del coseno

Lado del triángulo:

☒ Filtrado anisotrópico

Repeticiones:

Coeficientes de difusión:

Otros parámetros:

Nivel Rojo para líneas:

Nivel Verde para líneas:

Nivel Azul para líneas:

Número de objeto glotis:

☒ Individualiza glotis en la imagen.

☐ Resalta glotis segmentada en la imagen.

☒ Guardar imagen segmentada.

Explorar

☒ Guardar imagen preprocesada

☐ Guarda máscara de objeto indicado como glotis.

Rendimiento:

Tiempo de operación:

Realizar segmentación

Figura 5.1: Pantalla principal de la aplicación.

La pantalla principal se ha dividido en 3 partes:

- Explorador de archivos: donde se deberá escoger el fichero a analizar.
- Selección de parámetros: donde el usuario escogerá dentro de una serie de parámetros de los que dependerá el posterior análisis.
- Rendimiento: donde se muestra el tiempo empleado en el análisis.

5.1.1. Explorador de archivos

Se corresponde con la parte superior de la ventana de la aplicación. La parte principal es un explorador de archivos que se abre pulsando el botón *Explorar*, quedando la ruta

de dicho archivo impresa en el área *Fichero*:

La aplicación está preparada para trabajar con vídeos o con imágenes, tal y como muestran los dos *Checks* de la parte superior izquierda. Por defecto se supone que se trabajará con imágenes, pero el usuario puede cambiar de tipo cuando desee marcando el *Check: VIDEO*.

El explorador de archivos está preparado para filtrar únicamente los tipos de archivos permitidos por la aplicación: BMP¹, JPEG², IMG - en el caso de imágenes - y AVI³ en el caso de vídeos.

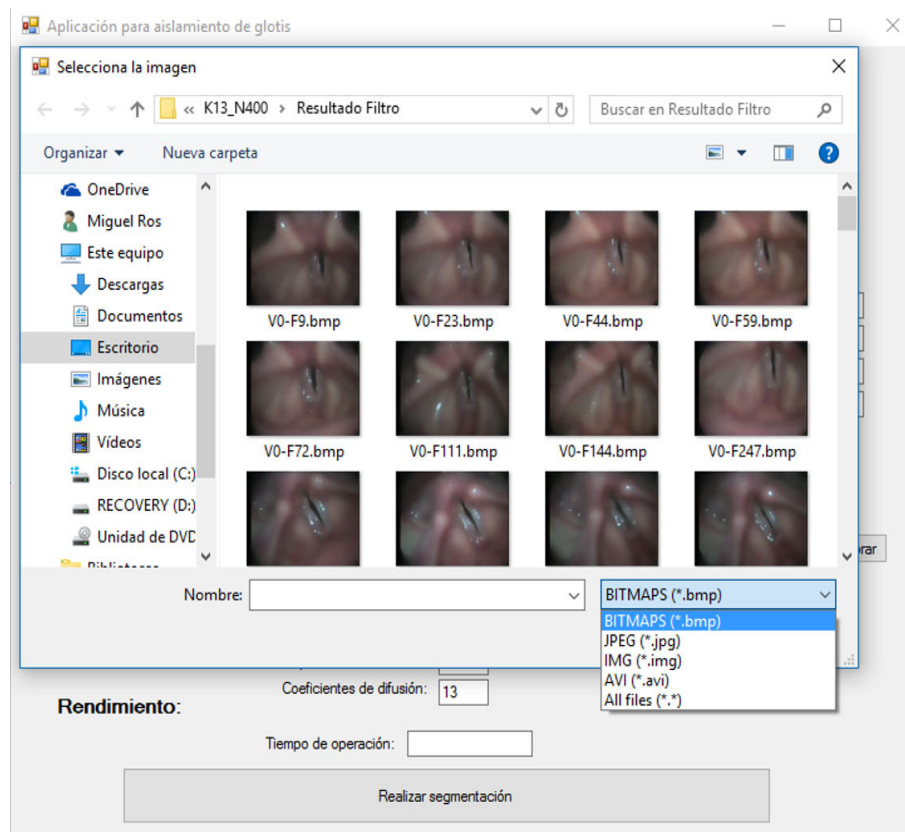


Figura 5.2: Ejemplo de selección de archivo.

¹Windows Bitmap

²Joint Photographic Experts Group

³Audio vídeo Interleave

En esta sección también hay un área para que, en caso de que el usuario seleccione un vídeo, pueda escoger distintos fotogramas. Debajo de dicho componente, se ha habilitado un *Label* indicando el número de fotogramas del vídeo para hacer más sencilla dicha tarea para el usuario. También se ha habilitado un *Check* que en caso de estar marcado, selecciona automáticamente todos los fotogramas del vídeo.

5.1.2. Selección de parámetros

En esta sección, el usuario deberá escoger entre una serie de parámetros. Se corresponde con la parte media de la pantalla, y como puede observarse, está dividida en 4 subsecciones, para hacer distinción de los parámetros a escoger.

1. Parámetros ajustables para la segmentación:

- a) Umbral de gradiente: Es el parámetro de umbralización del gradiente explicado en [2.5]. Por defecto tiene un valor de 2, pero en el caso de la segmentación de imágenes en color puede ser necesario cambiar su valor a 1.
- b) Porcentaje de área límite: Es el segundo factor de la ecuación [2.6] en el proceso de Merging basado en JND.
- c) Nivel de insensibilidad total: Parámetro que indica el umbral de sensibilidad del ojo humano utilizado en el proceso de Merging basado en JND.
- d) Mínimo de insensibilidad total: Parámetro que indica el mínimo valor a partir del cual se buscará el máximo umbral del histograma en caso de buscar dicho umbral en porcentaje.
- e) Umbral de coste para Merging: Parámetro que establece el umbral. Por encima del mismo se dificulta la unión entre regiones y por debajo lo facilita. Este es el parámetro principal de ajuste por el usuario.
- f) Tomar nivel de insensibilidad total en porcentaje: Marcando este *Check* indicamos que el parámetro *Mínimo de insensibilidad total* se toma en porcentaje.
- g) Aplicar Merging huérfanas de gris al finalizar: Marcando este *Check* indicamos que se realice el proceso de Merging de regiones huérfanas de gris al finalizar el proceso, explicado en [2.4.4]

2. Otros parámetros:

- a) Nivel de Rojo, Verde o Azul para líneas: Permite elegir el color de las líneas que se pintan en la imagen resultado. Estos 3 parámetros deben estar comprendidos entre 0 y 255, expresando el nivel de color en RGB.
- b) Individualiza glotis en la imagen y Número de objeto glotis: Una vez realizado el análisis, si el *Check: Individualiza glotis en la imagen* está marcado, aparecerá en *Número de objeto glotis* el número de objeto glotis detectado.

- c) Resalta glotis segmentada en la imagen: Marcando este *Check*, en la imagen resultado se observará el objeto glotis relleno con el color escogido por el usuario.
- d) Guardar imagen segmentada e imagen preprocesada: Estos *Checks* indican que se quiere guardar la imagen segmentada y/o la imagen preprocesada. Se ha habilitado un control en el que aparece la ruta predefinida donde se guardará la imagen resultado. La imagen preprocesada se guardará en un subdirectorio de esta ruta, diferenciando el tipo de preprocesado que se aplique. Ambas imágenes se guardan en formato *.BMP*
- e) Guardar máscara de objeto indicado como glotis: Marcando este *Check* además de guardar la imagen resultado, se guarda un nuevo archivo - *Máscara del objeto indicado como glotis* - en formato *.IMG*

3. Información de color:

- a) Aplicar Merging de color.
- b) Gradiente 3D.

Estos dos *Checks* permiten al usuario hacer un procesamiento de las imágenes atendiendo a la información de color, en lugar de hacerlo en escala de grises. Estas dos mejoras han sido programadas por el alumno Stepan Venzik.

4. Preprocesado:

- a) Transformada del coseno: Marcando este *Check* se indica que se desea hacer el preprocesado basado en la DCT. En caso de que el usuario desee utilizar este método, deberá escoger el lado del triángulo que desea aplicar, tal y como se explicó en el apartado [3.3.2]
- b) Filtrado anisotrópico: Marcando este *Check* se indica que se desea hacer el preprocesado basado en el filtrado anisotrópico. También, si el usuario opta por este análisis, deberá escoger dos parámetros: número de repeticiones del algoritmo y valor de los coeficientes de difusión. Ambos explicados en el apartado [3.3.3]

5.1.3. Rendimiento

Esta sección está situada en la parte inferior de la aplicación. Consta de un control *Tiempo de operación* - en el que aparece el tiempo destinado al análisis de la imagen elegida - y el botón que debe pulsar el usuario cuando haya configurado los parámetros. Una vez pulsado y realizado el análisis aparecerá una nueva pantalla que se explicará a continuación.

5.2. Pantalla resultados

La pantalla resultados se muestra a continuación:

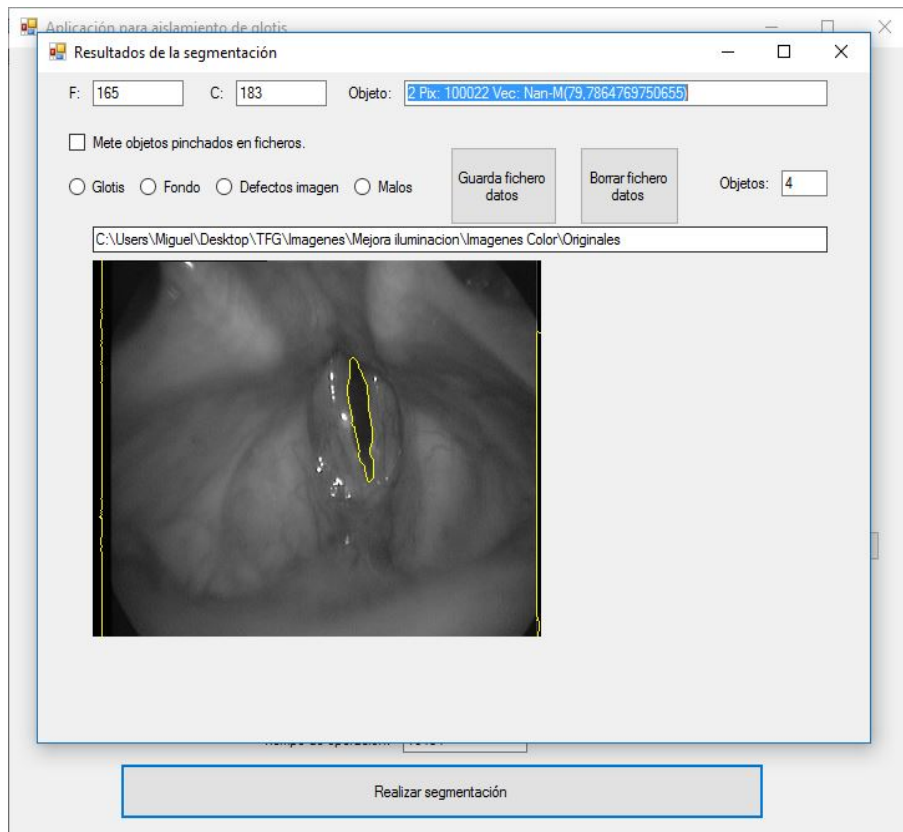


Figura 5.3: Pantalla resultados de la aplicación.

En la parte superior de la pantalla aparecen una serie de campos que se rellenarán en el caso de que el usuario “pinche” en alguna zona de la imagen final segmentada. F y C muestran el número de fila y de columna del pixel seleccionado respectivamente. Y el campo *Objeto:* muestra valores relativos al objeto seleccionado, como pueden ser: el número de objeto de entre todos los que componen la imagen, el número de píxeles total de la región o el valor medio del nivel de gris de la región.

Justo debajo de estos campos, aparecen una serie de marcadores mediante los cuales el usuario puede etiquetar el objeto seleccionado como: *Glotis*, *Fondo*, *Defectos imagen* o *Malos*. Así como guardar los parámetros de las regiones. Para poder hacer estas operaciones, el usuario debe seleccionar previamente una parte de la imagen.

En la parte inferior de la pantalla aparece una “caja” en la que se representa el resultado de la imagen. Esta imagen es en la que el usuario debe “pinchar” para seleccionar alguna de las regiones. Cabe destacar que la imagen que aparece en esta pantalla es la misma que se guarda en el directorio previamente comentado.

Encima de la imagen resultado, aparece un campo en el que se deja impreso el directorio de la imagen que se ha analizado. Si el usuario no cambia la ruta, será en un subdirectorio de esta ruta donde se guardarán los archivos generados.

Capítulo 6

Consideraciones finales

En este apartado se pretende dar una visión global del trabajo realizado, aportando las conclusiones obtenidas así como una serie de posibles líneas futuras para posteriores trabajos.

6.1. Conclusiones

Este trabajo tiene dos grandes objetivos, por un lado migrar la aplicación desarrollada en [1] y por el otro añadirle una serie de funcionalidades para tratar de mejorar la detección del área glotal.

En primer lugar, la migración de la aplicación tiene dos posibles conclusiones. Se ha conseguido migrar a un entorno más moderno, a un lenguaje mucho más versátil y que seguro facilitará el trabajo para posteriores mejoras. Por el contrario, al realizar la migración ha bajado notoriamente el rendimiento de la aplicación atendiendo a los tiempos de análisis. Esto se debe principalmente a que no se ha tenido en cuenta la gestión de memoria en este nuevo entorno.

En segundo lugar, en relación a las mejoras introducidas se pueden sacar varias conclusiones:

1. Local Normalization: La implementación del algoritmo LN - explicado en [3.3.1] - no consigue los resultados esperados. No se consigue la suficiente uniformidad de color como para detectar la glotis con mayor exactitud, en algunos casos hasta empeora los resultados obtenidos con la aplicación inicial.

2. Transformada discreta del coseno:

Mediante el algoritmo de la DCT - explicado en [3.3.2] - se consiguen mejores resultados que con el algoritmo anterior. Es cierto que centrándose en las medias sí que se consigue una mejora visible, pero analizando imagen por imagen la mejora no es tan evidente.

3. Filtro anisotrópico:

Aplicando este preprocesado - explicado en [3.3.3] - se consiguen unos resultados similares a los obtenidos aplicando la *DCT*. El problema de este tipo de algoritmo es el rendimiento, ya que el tiempo de preprocesado de cada imagen aumenta considerablemente.

De un modo general se puede concluir que este proyecto ha cumplido los objetivos planteados en su inicio y puede servir como base para plantear nuevas mejoras.

6.2. Líneas futuras

Una vez analizada cada parte de este proyecto, surgen varias ideas sobre las que basar nuevas propuestas de trabajo, ya que quedan muchas líneas abiertas para lograr que la aplicación funcione correctamente en todos los aspectos.

Algunos de estos aspectos podrían ser:

1. Analizar el nuevo entorno de trabajo pudiendo así realizar las mejoras oportunas a nivel de código para reducir así tiempos de análisis y mejorar con esto el rendimiento.
2. Implementar nuevos métodos de preprocesado de las imágenes.
3. Revisar las mejoras implementadas con el fin de mejorar los resultados analizando nuevos vídeos con distintos parámetros.
4. Hacer un estudio previo del vídeo o de la imagen a analizar y determinar así el tipo de preprocesado que mejores resultados puede obtener.

Capítulo 7

Anexos

Preprocesado para la mejora de la segmentación del área glotal.

	ORIGINAL	Tr=9	Tr=12	Tr=15	Tr=18
Imagen	Jaccard index				
V0-F144-S.img	0,876957494	0,461783439	0,840437158	0,860674157	0,860830527
V0-F23-S.img	0,877068558	0,862068966	0,863258026	0,864447087	0,851764706
V0-F247-S.img	0,929471033	0,924850555	0,927288281	0,926495726	0,922354949
V0-F44-S.img	0,861517976	0,818998716	0,802005013	0,735294118	0,844
V0-F59-S.img	0,912941176	0,90736342	0,906398104	0,599761051	0
V0-F72-S.img	0,916195857	0,902921772	0,902830189	0,904716981	0,871277618
V0-F9-S.img	0	0,894829097	0,893076249	0,719000892	0,903111111
MEDIA	0,767736013	0,824687995	0,876470431	0,801484288	0,750476987
V1-F118-S.img	0,901717557	0,741594621	0,771373679	0,85975024	0,816696915
V1-F137-S.img	0,873443983	0,840370752	0	0,845993757	0,843912591
V1-F142-S.img	0,874418605	0,831228473	0,831381733	0,833723653	0,831381733
V1-F26-S.img	0,893687708	0,857142857	0,798990749	0,856302521	0,856302521
V1-F48-S.img	0,871440897	0,832761578	0,8392702	0,839130435	0,83753258
V1-F73-S.img	0	0,850196078	0,726758286	0,87065481	0,868122977
V1-F76-S.img	0,88012959	0,31768559	0,825897715	0,824809576	0,824809576
V1-F95-S.img	0,889923793	0,864841745	0,866438356	0,866323907	0,866438356
MEDIA	0,773095267	0,766977712	0,70751384	0,849586112	0,843149656
V10-F0-S.img	0,859923826	0,90166205	0,913523459	0,916474389	0,912863071
V10-F22-S.img	0,809420916	0,236104279	0,429679922	0,218549581	0,39523578
V10-F26-S.img	0,830891586	0,803921569	0,813142602	0,40499554	0,808225302
V10-F4-S.img	0,855066922	0,903895004	0,74545295	0,917900973	0,913061917
V10-F48-S.img	0,862546816	0,857538583	0,856971212	0,848545237	0,848146672
V10-F51-S.img	0,833432481	0,121613486	0,724055189	0,131848284	0,15833835
V10-F6-S.img	0,457294029	0,188871473	0,458430458	0,822307692	0,36228482
V10-F69-S.img	0,571145265	0	0	0,81868969	0,323304437
MEDIA	0,75996523	0,501700806	0,63736868	0,634913923	0,590182544
V11-F0-S.img	0,796846011	0,801010101	0,80465587	0,80182002	0,801010101
V11-F14-S.img	0,834942932	0,756647865	0,74554295	0,792436226	0,792436226
V11-F27-S.img	0,351724138	0,767772512	0,77014218	0,332147094	0,772241993
V11-F32-S.img	0,757333333	0,719041278	0,721704394	0,719041278	0,71770972
V11-F45-S.img	0,897274633	0,862827225	0,83148335	0,848421053	0,853530032
V11-F58-S.img	0,812734082	0,695238095	0,740484429	0,785982478	0,788485607
V11-F75-S.img	0,855734767	0,745908028	0,766639277	0,824057451	0,823159785
V11-F90-S.img	0,883351007	0,860064586	0,861141012	0,807654563	0,860064586
MEDIA	0,773742613	0,776063711	0,780224183	0,738945022	0,801079757
V12-F124-S.img	0,835396941	0,293896006	0,845679012	0	0
V12-F127-S.img	0,021052632	0,800738007	0	0	0,798277983
V12-F141-S.img	0,830645161	0,727019499	0,793879388	0,800371402	0,333333333
V12-F15-S.img	0,850393701	0,835262689	0,826468011	0	0,833333333
V12-F33-S.img	0,808607022	0,79753915	0,767653759	0,673120729	0,673120729
V12-F48-S.img	0	0,806902219	0,830508475	0	0,842235004
V12-F64-S.img	0,906273063	0,892936803	0,8929491018	0,879491018	0,876304022
MEDIA	0,607481217	0,736327768	0,580598378	0,33614045	0,622372058
V13-F11-S.img	0	0,840544872	0,844534413	0,836171938	0,838576159
V13-F14-S.img	0	0,856213403	0	0,859016393	0,867509987
V13-F18-S.img	0,52972973	0,48317757	0	0,56	0,450413223
V13-F40-S.img	0	0,85452882	0,491162791	0,845940959	0,838112858
V13-F45-S.img	0	0,822368421	0	0,608255452	0
V13-F67-S.img	0,842248413	0,868421053	0,854788877	0,858024691	0,849639547
V13-F71-S.img	0,646106399	0	0,847758887	0,843364198	0,450310559
V13-F9-S.img	0,4	0,81610247	0,809698079	0,832855778	0,827420901
MEDIA	0,302260568	0,692669576	0,480992881	0,780453676	0,640247904
V14-F117-S.img	0,707943925	0,84107946	0,84939759	0,846385542	0,835082459
V14-F17-S.img	0,836023789	0,764551163	0,799069767	0,778393352	0,78597786
V14-F20-S.img	0	0,860257681	0,867850099	0,861928934	0
V14-F21-S.img	0,684500393	0,742619926	0,766666667	0,759172154	0,766381766
V14-F46-S.img	0	0	0,743639922	0,772123894	0,818083961
V14-F91-S.img	0,030478955	0,841435185	0,867058824	0,395092025	0,856976744
V14-F93-S.img	0,719879518	0	0,821243523	0,813648294	0,783375315
MEDIA	0,425546654	0,578577631	0,816418056	0,746677742	0,692268301
V2-F16-S.img	0	0,862152358	0,853864734	0	0
V2-F17-S.img	0,828244275	0,795112782	0,858178054	0,866315789	0,878074866
V2-F3-S.img	0,862793522	0,85977709	0	0,858926342	0,86267166
V2-F32-S.img	0,871201158	0	0	0,887864823	0,890769231
V2-F47-S.img	0,833040422	0,844827586	0,844827586	0,83908046	0,837164751
V2-F63-S.img	0,880269815	0	0,81878089	0,818481848	0
V2-F79-S.img	0,8374613	0,745269287	0,780152672	0	0
MEDIA	0,730430077	0,586662817	0,593868277	0,610095609	0,495525787
V3-F108-S.img	0,844856661	0,793570722	0,793570722	0,793570722	0,793570722
V3-F150-S.img	0,82788296	0,836158192	0,837429112	0,835538752	0,835538752
V3-F247-S.img	0,838669951	0	0,886486486	0,869623656	0,883378016
V3-F28-S.img	0,750439367	0,704174229	0,773547094	0,725252525	0,727272727
V3-F66-S.img	0,797368421	0,783519553	0,841317365	0,835347432	0,841317365
MEDIA	0,811843472	0,623484439	0,826470055	0,811866517	0,816215416
V4-F103-S.img	0	0,398	0,751491054	0,748514851	0
V4-F119-S.img	0	0,793977813	0,786686838	0,784829721	0,791344466
V4-F122-S.img	0	0,559454191	0,5546875	0,678456592	0,621705426
V4-F43-S.img	0	0,311455847	0	0,387692308	0,808788599
V4-F6-S.img	0	0,809664694	0,834132311	0,887475538	0,891625616
V4-F62-S.img	0,332896461	0	0,731736527	0,722933644	0,732941176
V4-F85-S.img	0,506507592	0,541509434	0,746924429	0,705667276	0,602564103
MEDIA	0,119914865	0,48772314	0,629379808	0,702224276	0,635567084
V5-F102-S.img	0,880579011	0,847304541	0,847304541	0,824039653	0,85235732
V5-F24-S.img	0,915338645	0,528992879	0,897461929	0,529170931	0,897897898
V5-F29-S.img	0,800210305	0,785106383	0,784877529	0,809151786	0,830275229
V5-F41-S.img	0,890468227	0,511489362	0,509787234	0,518932874	0,87649063
V5-F57-S.img	0,83580786	0,809815951	0,820713664	0	0
V5-F61-S.img	0,868556701	0,809210526	0,842618384	0,842618384	0,819783198
V5-F73-S.img	0,878600823	0	0,561833689	0	0,865546218
V5-F88-S.img	0,877486911	0,532700422	0,836122914	0,838266385	0,805111821
MEDIA	0,86838106	0,603088758	0,762674985	0,545272502	0,743432789
V6-F31-S.img	0,88478676	0,852323361	0,851050286	0,852323361	0,849140675
V6-F38-S.img	0,813909774	0,769818529	0,759469697	0,652131783	0,757516974
V6-F51-S.img	0,916222391	0,892530488	0,891768293	0,890243902	0
V6-F57-S.img	0,935006435	0,910737387	0,910090556	0,910737387	0,908796895
V6-F61-S.img	0,910377358	0,869606004	0,873814042	0,871916509	0,34662868
V6-F73-S.img	0,907815631	0,877755511	0,877755511	0,877755511	0,877755511
V6-F98-S.img	0,899047619	0,875476493	0,87778485	0,876590331	0,876590331
MEDIA	0,895309424	0,864035396	0,863104748	0,84738554	0,659489867
V7-F107-S.img	0,816935484	0,785185185	0,792845258	0,796157059	0,790309106
V7-F3-S.img	0,886855241	0,807071483	0,863070539	0,865384615	0,862155388
V7-F30-S.img	0,882867133	0,864840989	0	0,887465691	0
V7-F40-S.img	0,734770384	0	0,743396226	0,626865672	0,843396226
V7-F59-S.img	0	0	0	0,316910786	0,304
V7-F75-S.img	0,744797371	0,768133175	0,643281807	0,772241993	0,765755054
V7-F99-S.img	0,879166667	0,863869863	0,470537262	0,850862069	0,46875
MEDIA	0,706484611	0,584157242	0,50187587	0,730841126	0,576337968
V8-F0-S.img	0,841743119	0,778732546	0	0,512455516	0,798339265
V8-F13-S.img	0	0	0	0	0,266165414
V8-F16-S.img	0,778625954	0	0	0	0
V8-F29-S.img	0	0,778195489	0	0,740740741	0,783479349
V8-F47-S.img	0,660618997	0,80141844	0,811965812	0,829268293	0,837876614
V8-F64-S.img	0,748510131	0	0,814223512	0,779562044	0,823964497
V8-F78-S.img	0	0,75620438	0,817292007	0,78202995	0,828618968
MEDIA	0,432785457	0,444935836	0,349068762	0,520579506	0,61977773
V9-F34-S.img	0,818807339	0,804038005	0,80890973	0,801662708	0,802850356
V9-F36-S.img	0	0,879008746	0,847482014	0,852173913	0,851156099
V9-F38-S.img	0,724470135	0,764478764	0,767976424	0,659615385	0,665346535
V9-F4-S.img	0	0,799191375	0,805929919	0,799191375	0,805929919
V9-F65-S.img	0,869297164	0,858942065	0,862199747	0,85372005	0,858942065
V9-F69-S.img	0,684303351	0,608695652	0,673321234	0,674545455	0,661896243
V9-F79-S.img	0	0,859354839	0,863577864	0,862820513	0,859173127
V9-F84-S.img	0,773351648	0,783430233	0,715133531	0,712166172	0,721804511
MEDIA	0,483778705	0,79464246	0,780566308	0,776986946	0,778387353

Figura 7.1: Resultados DCT

	ORIGINAL	K=3 N=10	K=3 N=500	K=13 N=10	K=13 N=100	K=13 N=400
Imagen	Jaccard index					
V0-F144-S.img	0,876957494	0	0	0	0	0
V0-F23-S.img	0,877068558	0,877068558	0,789862724	0,863109049	0	0
V0-F247-S.img	0,929471033	0,935864979	0,846681922	0,921666667	0	0
V0-F44-S.img	0,861517976	0,852	0,73288136	0	0	0
V0-F59-S.img	0,912941176	0,915194346	0,849785408	0	0	0
V0-F72-S.img	0,916195857	0,914393227	0,860861759	0,911710037	0	0
V0-F9-S.img	0	0	0,916370107	0	0	0
MEDIA	0,767736013	0,642074444	0,714407151	0,38521225	0	0
V1-F118-S.img	0,901717557	0,902764538	0,772	0,898192198	0	0
V1-F137-S.img	0,873443983	0,857862903	0,868154158	0	0	0
V1-F142-S.img	0,874418605	0,875	0,844492441	0,8757259	0	0,801614764
V1-F26-S.img	0,893687708	0,89824854	0,902074689	0	0	0
V1-F48-S.img	0,871440897	0,870466321	0,825769854	0,881385281	0	0
V1-F73-S.img	0	0	0,903481013	0	0	0
V1-F76-S.img	0,88012959	0,881081081	0,896103896	0	0	0
V1-F95-S.img	0,889923793	0,891617273	0,891047297	0,896697714	0	0,865206403
MEDIA	0,773095267	0,772130082	0,862890418	0,444000137	0	0,208352646
V10-F0-S.img	0,859923826	0,851588629	0,501387604	0,876883341	0	0
V10-F22-S.img	0,809420916	0,409111013	0,402858419	0,425276753	0,838781908	0,843482183
V10-F26-S.img	0,830891586	0,83529418	0,839189748	0,811958763	0,858218875	0
V10-F4-S.img	0,855066922	0,861057507	0,8812	0,920131741	0	0
V10-F48-S.img	0,862546816	0,858361142	0,845259717	0,877714286	0	0
V10-F51-S.img	0,833432481	0,828402367	0,835545297	0,836624776	0	0
V10-F6-S.img	0,457294029	0,431784108	0,418419144	0,829512894	0	0
V10-F69-S.img	0,571145265	0	0,381305638	0,367418546	0,003900536	0,843175218
MEDIA	0,75996523	0,63444986	0,638145696	0,743190137	0,242985903	0,210832175
V11-F0-S.img	0,796846011	0,848756239	0,795158287	0,832847425	0	0
V11-F14-S.img	0,834942932	0,842844601	0,870659722	0,8615148862	0	0
V11-F27-S.img	0,351724138	0,355608592	0,817969662	0,344457688	0	0
V11-F32-S.img	0,757333333	0,752	0,765333333	0,764	0	0
V11-F45-S.img	0,897274633	0,894129979	0,888655462	0,902413431	0	0
V11-F58-S.img	0,812734082	0,823970037	0,836658354	0	0	0
V11-F75-S.img	0,855734767	0,855734767	0,741886792	0,856759176	0	0
V11-F90-S.img	0,883351007	0,87881981	0,877637131	0,886975242	0	0
MEDIA	0,773742613	0,781483001	0,824244843	0,681575228	0	0
V12-F124-S.img	0,835396941	0,854151085	0	0,846210996	0	0
V12-F127-S.img	0,021052632	0,856123662	0	0,14408867	0	0
V12-F141-S.img	0,830645161	0,829422383	0,827215756	0,830948122	0	0
V12-F15-S.img	0,850393701	0,852173913	0	0,858676208	0	0
V12-F33-S.img	0,808607022	0,807474519	0,8125	0	0	0
V12-F48-S.img	0	0,876190476	0,771609276	0,876886418	0	0
V12-F64-S.img	0,906273063	0,8099631	0	0,912946429	0,895400593	0
MEDIA	0,607481217	0,855071305	0,344475005	0,638536692	0,12791437	0
V13-F11-S.img	0	0	0,748322148	0,84566428	0,833596215	0,848582996
V13-F14-S.img	0	0	0	0,866878981	0,867132867	0,875884244
V13-F18-S.img	0,529727973	0	0,322757112	0,703111111	0,73458445	0,741613781
V13-F40-S.img	0	0	0	0	0,866666667	0,868677906
V13-F45-S.img	0	0	0,818978102	0,837293017	0	0
V13-F67-S.img	0,842248413	0,851102941	0,851376147	0,897187197	0,878672986	0
V13-F71-S.img	0,646106399	0	0	0,872563718	0,875094056	0,867362146
V13-F9-S.img	0,4	0	0,816901408	0,831541219	0,821808511	0
MEDIA	0,302260568	0,106387868	0,444791865	0,73177994	0,734694469	0,525265134
V14-F117-S.img	0,707943925	0,78768021	0,82231405	0,833568406	0,863235294	0,865300146
V14-F17-S.img	0,836023789	0,838765009	0,771781033	0,871287129	0	0
V14-F20-S.img	0	0,735641227	0,811201445	0,681519357	0	0
V14-F21-S.img	0,684500393	0,764134276	0,819458375	0,755420642	0	0
V14-F46-S.img	0	0	0,623898139	0	0	0
V14-F91-S.img	0,030478955	0,703075292	0	0,757281553	0	0
V14-F93-S.img	0,719879518	0,642473118	0	0,679156909	0	0
MEDIA	0,425546654	0,638824162	0,549807577	0,654033428	0,123319328	0,123614307
V2-F16-S.img	0	0	0,889428919	0	0	0
V2-F17-S.img	0,828244275	0,83253128	0,81988743	0	0	0
V2-F3-S.img	0,862793572	0,86509901	0,874529486	0,880394575	0,882425743	0,879801735
V2-F32-S.img	0,871201158	0,776349614	0	0,905405405	0	0
V2-F47-S.img	0,833040422	0,872897196	0	0,885057471	0	0
V2-F63-S.img	0,880269815	0,88047138	0,806990881	0,882352941	0	0
V2-F79-S.img	0,8374613	0	0,825617284	0,856240126	0	0
MEDIA	0,730430077	0,603906926	0,602350571	0,629921503	0,12606082	0,125685962
V3-F108-S.img	0,844856661	0,821003928	0,81350335	0,839802801	0,288851351	0,288851351
V3-F150-S.img	0,82788296	0,802325581	0,884191176	0,894934334	0,892655367	0,869402985
V3-F247-S.img	0,838669951	0,836385836	0,762897914	0,924119241	0,915662651	0
V3-F28-S.img	0,750439367	0,761384335	0,80608365	0,818359375	0	0
V3-F66-S.img	0,797368421	0,789336801	0,820689655	0,888888889	0,886056972	0,845468053
MEDIA	0,811843472	0,802207296	0,818733518	0,872640928	0,596645268	0,400744478
V4-F103-S.img	0	0,727758007	0	0,587771203	0,698924731	0,774859287
V4-F119-S.img	0	0,685847589	0,695583596	0,703296703	0,858914729	0,863772455
V4-F122-S.img	0	0,361198738	0,802197802	0,554702495	0,791512915	0
V4-F43-S.img	0	0	0	0	0,851605759	0
V4-F6-S.img	0	0	0	0,804549284	0,868778281	0,899530516
V4-F62-S.img	0,332896461	0	0	0,560759494	0,76119403	0
V4-F85-S.img	0,506507592	0	0	0,793286219	0,732441472	0,822183099
MEDIA	0,119934865	0,253543476	0,213968771	0,5720522	0,794767417	0,480049337
V5-F102-S.img	0,880579011	0,879854369	0,8848415962	0,8848415962	0	0
V5-F24-S.img	0,915338645	0,920920921	0,903441683	0,933200399	0	0
V5-F29-S.img	0,800210305	0,847112118	0,815843621	0,840757238	0	0
V5-F41-S.img	0,890468227	0,904040404	0,909704641	0,901098901	0	0
V5-F57-S.img	0,83580786	0,836681223	0,736048265	0,829732066	0	0
V5-F61-S.img	0,868556701	0,863753213	0,848180678	0,861715749	0	0
V5-F73-S.img	0,878600823	0,880412371	0,803113553	0,890269151	0	0
V5-F88-S.img	0,877486911	0,878378378	0,871007684	0,88247639	0	0
MEDIA	0,86838106	0,876394125	0,846989372	0,878551182	0	0
V6-F31-S.img	0,88478676	0,885423297	0,881604074	0,88478676	0,885568976	0
V6-F38-S.img	0,813909774	0,662463628	0,652764306	0,828408008	0,832535885	0,833969466
V6-F51-S.img	0,916222391	0,916222391	0,911077619	0,894977169	0	0
V6-F57-S.img	0,935006435	0,935649936	0,935732648	0,935649936	0,938946015	0
V6-F61-S.img	0,910377358	0,906603774	0,86983842	0,910546139	0,879706152	0
V6-F73-S.img	0,907815631	0,893596059	0,907278166	0,909090991	0,902195609	0,903103613
V6-F98-S.img	0,899047619	0,899494311	0,8988622	0,903492063	0	0
MEDIA	0,895309424	0,871350485	0,865308205	0,895395712	0,634136091	0,248166154
V7-F107-S.img	0,816935484	0,833744856	0,816260163	0,836228288	0	0
V7-F3-S.img	0,886855241	0,8875	0,807458143	0,894648829	0	0
V7-F30-S.img	0,882867133	0,883944154	0,712891986	0,922445255	0	0
V7-F40-S.img	0,734770384	0,879924953	0,841601393	0	0	0
V7-F59-S.img	0	0	0	0,328457447	0	0
V7-F75-S.img	0,744797371	0,779605263	0,785545024	0,816132859	0	0
V7-F99-S.img	0,879166667	0,890862944	0,890987124	0,889649273	0	0
MEDIA	0,706484611	0,736511739	0,693534833	0,669651707	0	0
V8-F0-S.img	0,841743119	0,78179938	0,826985854	0,827922078	0	0
V8-F13-S.img	0	0	0,653088042	0,691460055	0,716524217	0
V8-F16-S.img	0,778625954	0,789415656	0,725469729	0,794150731	0	0
V8-F20-S.img	0	0	0,796983759	0,842302879	0,829177057	0
V8-F47-S.img	0,660618997	0,691371681	0,829201102	0,854014599	0	0
V8-F64-S.img	0,748510131	0	0,819736842	0,866764275	0	0
V8-F78-S.img	0	0,805766313	0,791291291	0,85785124	0,860883797	0,845257903
MEDIA	0,432785457	0,438336147	0,77753666	0,819209408	0,343797867	0,120751129
V9-F34-S.img	0,818807339	0,823529412	0	0,827626919	0	0
V9-F36-S.img	0	0,81938911	0,840929874	0,847977685	0	0
V9-F38-S.img	0,724470135	0,618955513	0,71969697	0,705653021	0	0
V9-F4-S.img	0	0	0,829105474	0,845222073	0,829986613	0
V9-F65-S.img	0,869297164	0,866171004	0,865763547	0,884184309	0	0
V9-F69-S.img	0,684303351	0,636363636	0	0	0	0
V9-F79-S.img	0	0	0,816091954	0,882205514	0,870927318	0
V9-F84-S.img	0,773351648	0,762872629	0,702857143	0,727665706	0	0
MEDIA	0,483778705	0,565910163	0,59793062	0,715066903	0,212614241	0

Figura 7.2: Resultados Filtro

Preprocesado para la mejora de la segmentación del área glotal.

Imagen	ORIGINAL	DCT	FILTRO	DCT + FILTRO
Vo-F144-S.img	0,876057494	0,840437158	0	0,87804878
Vo-F23-S.img	0,877068558	0,863258026	0,863109049	0,863744076
Vo-F247-S.img	0,929471033	0,927288281	0,921666667	0,921967769
Vo-F44-S.img	0,861517976	0,802005013	0	0
Vo-F59-S.img	0,912941176	0,906398104	0	0
Vo-F72-S.img	0,916195857	0,902830189	0,911710037	0,91020794
Vo-F9-S.img	0	0,893076249	0	0
MEDIA	0,767736013	0,876470431	0,38521225	0,510566938
V1-F118-S.img	0,901717557	0,771373679	0,898192198	0,867689358
V1-F137-S.img	0,873443983	0	0	0,837090164
V1-F142-S.img	0,874418605	0,831381733	0,8757259	0,864644107
V1-F26-S.img	0,893687708	0,798990749	0	0,864207879
V1-F48-S.img	0,871440897	0,8392702	0,881385281	0,846220678
V1-F73-S.img	0	0,726758286	0	0,86742118
V1-F76-S.img	0,88012959	0,825807715	0	0,838953386
V1-F95-S.img	0,889923793	0,866438356	0,896697714	0,877777778
MEDIA	0,773095267	0,70751384	0,444000137	0,858000816
V10-F0-S.img	0,859923826	0,913523459	0,876883341	0,916361974
V10-F22-S.img	0,809420916	0,429679922	0,425276753	0,857210402
V10-F26-S.img	0,830891586	0,813142602	0,811958763	0,820295038
V10-F4-S.img	0,855066922	0,903145695	0,920313741	0,92323914
V10-F48-S.img	0,862546816	0,856972112	0,877714286	0,86859865
V10-F51-S.img	0,833432481	0,724055189	0,836624776	0,374473209
V10-F6-S.img	0,457294029	0,458430458	0,829512894	0,367362722
V10-F69-S.img	0,571145265	0	0,344457688	0,327115409
MEDIA	0,75996523	0,63736868	0,743190137	0,681836931
V11-F0-S.img	0,796846011	0,80465587	0,832847425	0,828600406
V11-F14-S.img	0,834942932	0,74554295	0,865148862	0,806848112
V11-F27-S.img	0,351721138	0,77014218	0,344457688	0,776859504
V11-F32-S.img	0,757333333	0,721704394	0,764	0,704
V11-F45-S.img	0,897274633	0,83148335	0,902413431	0,874868559
V11-F58-S.img	0,812734082	0,740484429	0	0,796758105
V11-F75-S.img	0,855734767	0,766639277	0,856759176	0,826164875
V11-F90-S.img	0,883351007	0,861141012	0,886975242	0,860021209
MEDIA	0,773742613	0,780224183	0,681575228	0,809265096
V12-F124-S.img	0,835396941	0,845679012	0,846210996	0,848228043
V12-F127-S.img	0,021052632	0	0,14408867	0,831488315
V12-F141-S.img	0,830645161	0,793879388	0,830948122	0,818012999
V12-F15-S.img	0,850393701	0,826468011	0,858676208	0,829443447
V12-F33-S.img	0,808607022	0,767653759	0	0,781071836
V12-F48-S.img	0	0,830508475	0,876886418	0,843624609
V12-F64-S.img	0,906273063	0	0,912946429	0,880239521
MEDIA	0,607481217	0,580598378	0,638536692	0,833158409
V13-F11-S.img	0	0,844534413	0,84566428	0,835218094
V13-F14-S.img	0	0	0,866878981	0,86431833
V13-F18-S.img	0,52972973	0	0,703111111	0,733658537
V13-F40-S.img	0	0,491162791	0	0
V13-F45-S.img	0	0	0,837293017	0,847826087
V13-F67-S.img	0,842248413	0,854788877	0,897187197	0,879482072
V13-F71-S.img	0,646106399	0,847758887	0,872563718	0,838610039
V13-F9-S.img	0,4	0,809698079	0,831541219	0,818926297
MEDIA	0,302260568	0,480992881	0,73177994	0,727254932
V14-F117-S.img	0,707943925	0,84939759	0,833568406	0,841554559
V14-F17-S.img	0,836023789	0,799069767	0,871287129	0
V14-F20-S.img	0	0,867850099	0,681519357	0,887887888
V14-F21-S.img	0,684500393	0,766666667	0,755420642	0,783109405
V14-F46-S.img	0	0,743639922	0	0
V14-F91-S.img	0,030478955	0,867058824	0,757281553	0,448655257
V14-F93-S.img	0,719879518	0,821243523	0,679156909	0,825581395
MEDIA	0,425546654	0,816418056	0,654033428	0,540699786
V2-F16-S.img	0	0,853864734	0	0,869352869
V2-F17-S.img	0,828244275	0,858178054	0	0,868062827
V2-F3-S.img	0,862793572	0	0,880394575	0,86375
V2-F32-S.img	0,871201158	0	0,905405405	0,894333844
V2-F47-S.img	0,833040422	0,844827586	0,885057471	0,844827586
V2-F63-S.img	0,880260815	0,81878089	0,882351294	0,848633945
V2-F79-S.img	0,8374613	0,780152672	0,856240126	0,689655172
MEDIA	0,730430077	0,593686277	0,629921503	0,839803108
V3-F108-S.img	0,844856661	0,79357022	0,836902801	0,81556836
V3-F150-S.img	0,82788296	0,837429112	0,894934334	0,848771267
V3-F247-S.img	0,838669951	0,886486486	0,924119241	0,882591093
V3-F28-S.img	0,750439367	0,773547094	0,818359375	0,737373737
V3-F66-S.img	0,797368421	0,841317365	0,888888889	0,8444879518
MEDIA	0,811843472	0,826470055	0,872640928	0,82583649
V4-F103-S.img	0	0,751491054	0,587771203	0,683937824
V4-F119-S.img	0	0,786686838	0,703296703	0
V4-F122-S.img	0	0,5546875	0,554702495	0,772994129
V4-F43-S.img	0	0	0	0,860849057
V4-F6-S.img	0	0,834132311	0,804549284	0,890384615
V4-F62-S.img	0,332896461	0,731736527	0,560759494	0,736111111
V4-F85-S.img	0,506057592	0,746924429	0,793286219	0,725
MEDIA	0,119914865	0,629379808	0,5720522	0,667039534
V5-F102-S.img	0,880579011	0,847304541	0,889159562	0,855555556
V5-F24-S.img	0,915338645	0,897461929	0,933200399	0,919387755
V5-F29-S.img	0,800210305	0,784877529	0,840757238	0,826834862
V5-F41-S.img	0,890468227	0,509787234	0,901098901	0,876606684
V5-F57-S.img	0,83580786	0,820713664	0,829732066	0,822898032
V5-F61-S.img	0,868556701	0,842618384	0,861715749	0,865094003
V5-F73-S.img	0,878600823	0,561833689	0,890269151	0,876595745
V5-F88-S.img	0,877486911	0,836712914	0,88247639	0,852008457
MEDIA	0,86838106	0,762674985	0,878551182	0,861872187
V6-F31-S.img	0,88478676	0,851050286	0,88478676	0,855506047
V6-F38-S.img	0,813090774	0,759469697	0,828480008	0,617846751
V6-F51-S.img	0,916222391	0,891768293	0,894977169	0,893292683
V6-F57-S.img	0,935006435	0,910090556	0,935649936	0,920542636
V6-F61-S.img	0,910377358	0,873814042	0,910546139	0,873933649
V6-F73-S.img	0,907815631	0,877755511	0,90990991	0,882765531
V6-F98-S.img	0,899047619	0,87778485	0,903492063	0,881226054
MEDIA	0,895309424	0,863104748	0,895395712	0,846444764
V7-F107-S.img	0,816935484	0,792845258	0,836228288	0,801169591
V7-F3-S.img	0,886855241	0,863070539	0,894648829	0,820472441
V7-F30-S.img	0,882867133	0	0,922445255	0,90018315
V7-F40-S.img	0,734770384	0,743396226	0	0,866163996
V7-F59-S.img	0	0	0,328457447	0,318666667
V7-F75-S.img	0,744797371	0,643281807	0,816132859	0,781212842
V7-F99-S.img	0,879166667	0,470537262	0,889649273	0,871551724
MEDIA	0,706484611	0,50187587	0,669651707	0,765631487
V8-F0-S.img	0,841743119	0	0,827922078	0,803748622
V8-F13-S.img	0	0	0,691460055	0,683526012
V8-F16-S.img	0,778625954	0	0,794150731	0,78436019
V8-F29-S.img	0	0	0,842302879	0,80375
V8-F47-S.img	0,660618997	0,811965812	0,854014599	0,831845238
V8-F64-S.img	0,748510131	0,814223512	0,866764275	0,80952381
V8-F78-S.img	0	0,817292007	0,85785124	0,825291181
MEDIA	0,432785457	0,349068762	0,819209408	0,791720722
V9-F34-S.img	0,818807339	0,80890973	0,827626919	0,805164319
V9-F36-S.img	0	0,847482014	0,847977685	0,863570392
V9-F38-S.img	0,724470135	0,667976424	0,705653021	0,693069307
V9-F4-S.img	0	0,805929919	0,845222073	0,811320755
V9-F65-S.img	0,869297164	0,862199747	0,884184309	0,862373737
V9-F69-S.img	0,684303351	0,673321234	0	0,730072464
V9-F79-S.img	0	0,863577864	0,882205514	0,865806452
V9-F84-S.img	0,773351648	0,715135331	0,727665706	0,737704918
MEDIA	0,483778705	0,780566308	0,715066903	0,796135293

Figura 7.3: Resultados Filtro + DCT

Bibliografía

- [1] Osma Ruiz, V. J., “Contribución al procesado digital de imágenes para la caracterización de patologías laríngeas” Tesis doctoral, Universidad Politécnica de Madrid, Madrid, 2010.
- [2] Arthur C. Guiton, “Tratado de fisiología médica”, Ed. Interamericana, Quinta edición, 1976.
- [3] H Rouviere, “Anatomía humana. Descriptiva, topográfica y funcional”, Ed. Bailly-Bailliere, 1980.
- [4] FinK B.R, “The human larynx: a funtinal study”, New York, Ed. Raven press, 1974.
- [5] Portmann M. “Manual de Otorrinolaringología”, Ed. Toray-Masson, 1980.
- [6] Ramirez Camacho, R , “Manual de otorrinolaringología”, Ed. Mc Graw Hill, 2007.
- [7] Willian K. Pratt “Digital Image Processing” , *PIKS Scientific Inside*, cap. 8.
- [8] Caludio A. Perez, “Resumen 5” EL7007 Introducción al Procesamiento Digital de Imágenes , Departamento de Ingeniería Eléctrica, Universidad de Chile, 2014.
- [9] Estroboscopia, *MedWave*,[última consulta: 5 jul. 2016] Disponible en: [http :
//www.medwave.cl/link.cgi/Medwave/PuestaDia/Cursos/3083](http://www.medwave.cl/link.cgi/Medwave/PuestaDia/Cursos/3083)
- [10] Videolaringoestroboscopia, *Stel*,[última consulta: 5 jul. 2016] Disponible en: [http :
//stel.ub.edu/labfon/sites/default/files/XVII-24.pdf](http://stel.ub.edu/labfon/sites/default/files/XVII-24.pdf)
- [11] Laringoscopia, *MedlinePlus*,[última consulta: 5 jul. 2016] Disponible en: [https :
//www.nlm.nih.gov/medlineplus/spanish/ency/article/007507.htm](https://www.nlm.nih.gov/medlineplus/spanish/ency/article/007507.htm)
- [12] Programación orientada a objetos, *Desarrollo web*,[última consulta: 5 jul. 2016] Disponible en: [http :
//www.desarrolloweb.com/articulos/499.php](http://www.desarrolloweb.com/articulos/499.php)
- [13] Intervalo de confianza, *Support Office*,[última consulta: 5 jul. 2016] Disponible en: [https :
//support.office.com/es-es/article/Funcion-INTERVALO-CONFIANZA-75ccc007-f77c-4343-bc14-673642091ad6](https://support.office.com/es-es/article/Funcion-INTERVALO-CONFIANZA-75ccc007-f77c-4343-bc14-673642091ad6)

- [14] Fastcall, *MSDN*, [última consulta: 5 jul. 2016] Disponible en: [https :
//msdn.microsoft.com/es – es/library/6xa169sk.aspx](https://msdn.microsoft.com/es-es/library/6xa169sk.aspx)
- [15] CSharp, *MSDN*, [última consulta: 5 jul. 2016] Disponible en: [https :
//msdn.microsoft.com/es – es/library/kx37x362.aspx](https://msdn.microsoft.com/es-es/library/kx37x362.aspx)
- [16] Jaccard index, *Wikipedia, la enciclopedia libre*, [última consulta: 5 jul. 2016] Disponible en: [https : //es.wikipedia.org/wiki/Indice_Jaccard](https://es.wikipedia.org/wiki/Indice_Jaccard)
- [17] YCbCr, *Wikipedia, la enciclopedia libre*, [última consulta: 5 jul. 2016] Disponible en: [https : //en.wikipedia.org/wiki/YCbCr](https://en.wikipedia.org/wiki/YCbCr)
- [18] Borland, *Wikipedia, la enciclopedia libre*, [última consulta: 5 jul. 2016] Disponible en: [https : //es.wikipedia.org/wiki/Borland](https://es.wikipedia.org/wiki/Borland)
- [19] Filtrado anisotrópico, *Wikipedia, la enciclopedia libre*, [última consulta: 5 jul. 2016] Disponible en: [https : //es.wikipedia.org/wiki/Filtrado_anisotropico](https://es.wikipedia.org/wiki/Filtrado_anisotropico)
- [20] Fonética articulatoria, *Páginas personales*, [última consulta: 5 jul. 2016] Disponible en: [http : //paginaspersonales.deusto.es/airibar/Fonetica/Apuntes/02.html](http://paginaspersonales.deusto.es/airibar/Fonetica/Apuntes/02.html)
- [21] Anatomía laríngea, *Escuela médica*, [última consulta: 5 jul. 2016] Disponible en: [http :
//escuela.med.puc.cl/paginas/publicaciones/apuntesotorrino/AnatomiaLaringea.html](http://escuela.med.puc.cl/paginas/publicaciones/apuntesotorrino/AnatomiaLaringea.html)
- [22] Musculatura, *Mayullantalla blogspot*, [última consulta: 5 jul. 2016] Disponible en: [http : //mayullantalla.blogspot.com.es/2011/01/musculatura.html](http://mayullantalla.blogspot.com.es/2011/01/musculatura.html)
- [23] Vocal sound production, *hyperphysics*, [última consulta: 5 jul. 2016] Disponible en: [http : //hyperphysics.phy – astr.gsu.edu/hbasees/music/voice.html](http://hyperphysics.phy-astr.gsu.edu/hbasees/music/voice.html)
- [24] Patologías de la voz, *Clínica de vértigo*, [última consulta: 5 jul. 2016] Disponible en: [http : //clinicadevertigo.com/patologias – de – la – voz/](http://clinicadevertigo.com/patologias-de-la-voz/)